

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 28, 2005, 08:02:56 ; Search time 2756 Seconds

(without alignment)
10003.292 Million cell updates/sec

Title: US-10-600-230-1

Sequence: 1 gcgcgcgcgaaggaaggaagt.....cctccaatgatgaggtacc 485

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GeneBld:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sta:*
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: gb_hlg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|---------------|--------------------|
| 1 | 474.8 | 97.9 | 4046 | 15 SCAL1 | X57300 Saccharomyc |
| 2 | 474.8 | 97.9 | 33117 | 15 SCAL1 | X76078 S.cerevisia |
| 3 | 446.4 | 92.0 | 1089 | 15 SCYBR022W | Z35891 S.cerevisia |
| 4 | 287.4 | 59.3 | 534 | 15 AY557705 | AY557705 Saccharom |
| 5 | 284.2 | 58.6 | 4345 | 15 YSCPCHSY | M73697 S.cerevisia |
| 6 | 190.4 | 39.3 | 5176 | 15 SCYBR023C | Z35892 S.cerevisia |
| 7 | 63.2 | 13.0 | 102101 | 15 CR380948_4 | Continuation (5 of |
| 8 | 52.2 | 10.6 | 348174 | 2 CR382339 | CR382339 Plasmodi |
| 9 | 51.4 | 10.6 | 633 | 10 BV061001 | BV061001 S212P6318 |
| 10 | 51.4 | 10.6 | 158785 | 8 AC079954 | AC079954 Homo sapi |
| 11 | 51.4 | 10.6 | 174874 | 9 AC123054 | AC123054 Mus muscu |
| 12 | 51.4 | 10.6 | 206926 | 9 AC124497 | AC124497 Mus muscu |
| 13 | 51.2 | 10.6 | 196859 | 8 AC087277 | AC087277 Homo sapi |
| 14 | 50.4 | 10.4 | 193849 | 14 AC124696 | AC124696 Mus muscu |
| 15 | 50 | 10.3 | 244638 | 14 AC163463 | AC163463 Bos tauru |
| 16 | 48.8 | 10.1 | 168825 | 14 CR854985 | CR854985 Danto rer |
| 17 | 48.4 | 10.0 | 230657 | 14 AC095377 | AC095377 Rattus no |
| 18 | 48 | 9.9 | 181060 | 5 BX119907 | BX119907 Zebrafish |

| | | | | | |
|----|------|-----|--------|-------------|--------------------|
| 19 | 47.4 | 9.8 | 270629 | 14 CR354537 | CR354537 Danto rer |
| 20 | 47.2 | 9.7 | 1017 | 10 CNS06K1D | AL402311 T3 end of |
| 21 | 47.2 | 9.7 | 181223 | 14 CR376799 | CR376799 Danto rer |
| 22 | 47 | 9.7 | 1522 | 2 AY701231 | AY701231 Orconecte |
| 23 | 47 | 9.7 | 199684 | 8 AC005036 | AC005036 Homo sapi |
| 24 | 46.6 | 9.6 | 145925 | 5 BX842571 | BX842571 Zebrafish |
| 25 | 46.6 | 9.6 | 160339 | 9 AC132600 | AC132600 Mus muscu |
| 26 | 46.6 | 9.6 | 191707 | 14 CR847971 | CR847971 Danto rer |
| 27 | 46.4 | 9.6 | 120481 | 8 AC099738 | AC099738 Homo sapi |
| 28 | 46.4 | 9.6 | 217436 | 14 CR456634 | CR456634 Macropus |
| 29 | 46.2 | 9.5 | 5739 | 6 AX345620 | AX345620 Sequence |
| 30 | 46.2 | 9.5 | 6092 | 6 AX251775 | AX251775 Sequence |
| 31 | 46 | 9.5 | 180450 | 2 AB014835 | AB014835 Plasmodi |
| 32 | 46 | 9.5 | 254436 | 2 AB014827 | AB014827 Plasmodi |
| 33 | 46 | 9.5 | 330050 | 2 PF929355 | PF929355 Plasmodi |
| 34 | 45.6 | 9.4 | 145789 | 5 BX247871 | BX247871 Zebrafish |
| 35 | 45.6 | 9.4 | 168940 | 9 AC122434 | AC122434 Mus muscu |
| 36 | 45.6 | 9.4 | 207128 | 9 AC155816 | AC155816 Mus muscu |
| 37 | 45.4 | 9.4 | 191968 | 14 AC158414 | AC158414 Pongo pyg |
| 38 | 45.4 | 9.4 | 195351 | 5 AL928920 | AL928920 Zebrafish |
| 39 | 45.4 | 9.4 | 229302 | 14 AC023534 | AC023534 Homo sapi |
| 40 | 45.2 | 9.3 | 181168 | 5 BX293540 | BX293540 Zebrafish |
| 41 | 45.2 | 9.3 | 183141 | 14 CR407564 | CR407564 Danto rer |
| 42 | 45.2 | 9.3 | 200942 | 14 CR383681 | CR383681 Danto rer |
| 43 | 45 | 9.3 | 37515 | 6 AX58482 | AX58482 Sequence |
| 44 | 45 | 9.3 | 219084 | 14 CR749178 | CR749178 Danto rer |
| 45 | 45 | 9.3 | 223816 | 5 BX537162 | BX537162 Zebrafish |

ALIGNMENTS

| | | | | | |
|------------|--|---------|-----|--------|-----------------|
| RESULT 1 | SCAL1 | 4046 bp | DNA | linear | PLN 02-AUG-1991 |
| LOCUS | Saccharomycs cerevisiae CAL1 gene for chitin synthase 3. | | | | |
| DEFINITION | X57300.1 GI:3359 | | | | |
| ACCESSION | X57300 | | | | |
| VERSION | CAL1 gene; chitin synthase 3. | | | | |
| KEYWORDS | Saccharomycs cerevisiae (baker's yeast) | | | | |
| SOURCE | Saccharomycs cerevisiae | | | | |
| ORGANISM | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycs. | | | | |
| REFERENCE | 1 (bases 1 to 4046) | | | | |
| AUTHORS | Valdivieso,M.H., Mol,P.C., Shaw,J.A., Cabib,E. and Duran,A. | | | | |
| TITLE | CAL1, a gene required for activity of chitin synthase 3 in Saccharomycs cerevisiae | | | | |
| JOURNAL | J. Cell Biol. 114 (1), 101-109 (1991) | | | | |
| PubMed | 2050737 | | | | |
| REFERENCE | 2 (bases 1 to 4046) | | | | |
| AUTHORS | Duran,A. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (16-JAN-1991) A. Duran, Instituto de Microbiologia Bioquimica, Facultad de Biologia, CSIC/Universidad de Salamanca, 37008 Salamanca, Spain | | | | |
| FEATURES | Location/Qualifiers | | | | |
| source | 1..4046 | | | | |
| | /organism="Saccharomycs cerevisiae" | | | | |
| | /mol_type="genomic DNA" | | | | |
| | /db_xref="taxon:4932" | | | | |
| | /chromosome="II" | | | | |
| | /haplotype="partial Sau3a digest in Ycp50 vector" | | | | |
| | 240..244 | | | | |
| promoter | /note="PATA-like box" | | | | |
| | 253..257 | | | | |
| promoter | /note="PATA-like box" | | | | |
| | 279..3578 | | | | |
| gene | /gene="CAL1" | | | | |
| | 279..3578 | | | | |
| CDS | /gene="CAL1" | | | | |
| | /BC number="2.4.1.16" | | | | |
| | /codon_start=1 | | | | |
| | /evidence=experimental | | | | |

/product="chitin synthase"
/protein_id="CAA40559.1"
/db_xref="GI:3360"
/db_xref="GOA:P29465"
/db_xref="UniProt/Swiss-Prot:P29465"
/translation="MNHLDIPLSSGVPNPNRRSGSLRSKGSVSRKPSGRBETSYL
ODMTTDDKASVKISDEGVADPEPDGDVDNPESSSTOPINKSKIPRKRENDTSLF
WOMYCFITFMAPAPILACGMPKPKERQMAEKVALISVLIYALVAFITFETKT
VCSSTKRLKNNEVSLFVINGKAYELDTSRSGLIDPEVSDTLXPMSADGAS
PLFOVNGNCNHLITPKSNSIPHDDNNLAWPFCCKLNODSKSPFVENVYGMN
CHTSKEDRAFGKSKADVPTMDIKNSNMLIVYNGVDLIDLMLEEDVDVAP
VVFDDLKTSNLOQVDLSVLSNGHERKTLARCSLITIKGVDSKTVGCIASVLYVS
LVFILTSLVITIKETIACYRPTWYARQKATIVYNNKMDTENDIEMSNNSIOKVALKE
VDHLRPKKISKCSLGHKASTFDLLKQHSKMFQNSVIDLIDTSMSSLDSSGSTR
GMTTWTQVAMKLSNENKAVHSRNPSTLLPTSSMNKNTATSPVGSLSLOLSTII
HEDIVQOPLDPMYGFPLIHTICVTCSEDEBEGRTLLDLSSTHMDPKSLIMV
CDGLKSGNDKTPPEIALGMDDPTPEDEKPYSAVAVASGRHNMAKIYAFYK
YDSTIIPENQORPILITIVKCGTPABOGAARGRGRDQIILMSPLKRTPEERM
TQLEFOLKNTIQTIGLADPFEYTLMDADTKVPDALTTHVAMVADPLIMGCGE
TKIANKQSWTAIOVFETYSIHQAKAESVFGSVTLPGCFSMYRIKSPKSGDGM
VPLANKEVLSQRERMINSTVNLFEVLIRDLGTFPSMQFVIGIELIGTMVPLA
ICETVITIFALVSKPTVITVLVLAIIIGLIGLIVITATMSYILMMVCYICALPI
NMFWLPSYAVWKEPDPFSGWDTRTIAGNKKADENGERDHSKIMQRTWBEREDIL
NRKESDSFVA"

ORIGIN

Query Match 97.9%; Score 474.8; DB 15; Length 4046;
Best Local Similarity 99.6%; Pred. No. 3.4e-80;
Matches 476; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 CGAAGAGAGAAAGTGACTCTCTCGTGGGAGACAGTATGAATAATTTTACCTGTAT 65
DB 3546 CGAAGAGAGAAAGTGACTCTCTCGTGGAGATGACAGTATGAATAATTTTACCTGTAT 3605
QY 66 ACTTCAAGTATATATATGTTGTGTATATCTTATTTTGAAGGATTTTAAACACAC 125
DB 3606 ACTTCAAGTATATATATGTTGTGTATATCTTATTTTGAAGGATTTTAAACACAC 3665
QY 126 CTTAGAATAAATACTTAAATTAATTTCTCTATCTTTAAAGGACATTAACGTG 185
DB 3666 CTTAGAATAAATACTTAAATTAATTTCTCTATCTTTAAAGGACATTAACGTG 3725
QY 186 CTAGAGCAATACAGCTGATATCTGTAAACCTCATGTGGCCATTAATCTTCTAAC 245
DB 3726 CTAGAGCAATACAGCTGATATCTGTAAACCTCATGTGGCCATTAATCTTCTAAC 3785
QY 246 GGGTCTGTCTCTTCAAGGAGCTCCGAATATGCACTATTTATCTGTGGCATTTCCAA 305
DB 3786 GGGTCTGTCTCTTCAAGGAGCTCCGAATATGCACTATTTATCTGTGGCATTTCCAA 3845
QY 306 TTTATATCCCTTATGGGATTTGATGTGGCGTTTAAATAGTCAACGATGATCTTC 365
DB 3846 TTTATATCCCTTATGGGATTTGATGTGGCGTTTAAATAGTCAACGATGATCTTC 3905
QY 366 ACTTGTTCAGATTGCTCTTTGCTCTCTTAAAGTCTTCAATTTATCTTAAAGCAAGTT 425
DB 3906 ACTTGTTCAGATTGCTCTTTGCTCTCTTAAAGTCTTCAATTTATCTTAAAGCAAGTT 3965
QY 426 TGTATATCAAAATCTTGTCTCTTCTCCATGACCTTGAACCTCCAAAGAGAGAGTA 483
DB 3966 TGTATATCAAAATCTTGTCTCTTCTCCATGACCTTGAACCTCCAAAGAGAGAGTA 4023

RESULT 2
SCGAL1/c 33117 bp DNA 1linear PLN 18-APR-2005
LOCUS
DEFINITION S.cerevisiae (alpha288C) GAL1, FUR4 and CAL1 genes.
ACCESSION X76078
VERSION X76078.1 GI:498748
KEYWORDS CAL1 gene; chitin synthase 3; FUR4 gene; GAL1 gene; galactokinase;
ribosomal protein l28; uracil permease; YBR0304 gene; YBR0308 gene;
YBR0309 gene; YBR0310 gene; YBR0312 gene; YBR0313 gene; YBR0314

SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
gene
CDS

gene.
Saccharomyces cerevisiae (baker's yeast)
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
1 (bases 1 to 33117)
Smits,P.H.M., de Hann,M., Maat,C. and Grivell,L.A.
The complete sequence of a 33 kb fragment on the right arm of
chromosome II from Saccharomyces cerevisiae reveals 16 open reading
frames, five previously identified genes and a homologue of the
SCOL gene
Yeast 10, 75-80 (1994)
8091864
2 (bases 1 to 33117)
Smits,P.H.M.
Direct Submission
Submitted (11-NOV-1993) P.H.M. Smits, University of Amsterdam,
Section for Molecular Biology, Department of Molecular Cell
Biology, Kruislaan 318, 1098 SM Amsterdam, NETHERLANDS
Location/Qualifiers
1. .33117
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/strain="alpha288C"
/db_xref="taxon:4932"
2170. .3756
/gene="GAL1"
2170. .3756
/gene="GAL1"
/codon_start=1
/product="galactokinase"
/protein_id="CAA53677.1"
/db_xref="GI:498749"
/db_xref="GOA:P04385"
/db_xref="SGD:S000000224"
/db_xref="UniProt/Swiss-Prot:P04385"
/translation="WTKSHSEVYVPERNSAKELRPLRAKCSIIKKPISAYDAP
DFVAPSGKRVNLGSHIDYCDPSVLPALIDDMICAVYLNKPSITLINADPRAP
RKFDLPDGSVYITDPSVDSNFKGCLVHAFPLKCLAPERASALAGLQVFCG
DVPTGSGSSSAEFCALVALAVVAMGPGVHMSKONMLRTVAEAHVAGVNGMDQ
AASVCGEEDHALYVEFKPLKATPEFKPOLKHEISFVIANLVSNKFEAPNNYL
RVVEYTTAAVLAATYAGVLLSGKSGSTNGNLRDENVYVAYRHNISTPMNDIS
GIERLTRVLYVESILANKKQSFVDVYAOSSLNGSRPEFPEPDYLTTSPPROYKVQ
RAKHYTSESLVKAVKMLTTASTPADDFKQKQALMNEOASCDLKYESCCEIDK
ICSIALNSGYSGRITGAGWGCTVHLVPGGPNGITEKALANEFYKVPKITDA
ELENAIIVSKDALSCLYEL"
4592. .6493
/gene="FUR4"
4592. .6493
/gene="FUR4"
/codon_start=1
/product="uracil permease"
/protein_id="CAA53678.1"
/db_xref="GI:498750"
/db_xref="GOA:P05316"
/db_xref="interpro:IPR001248"
/db_xref="SGD:S000000225"
/db_xref="UniProt/Swiss-Prot:P05316"
/translation="WPDNLISLISGSSSRKLSRSLRQLESSENETFAPNNVDLEKEYSSQ
SNITTEVYASSEFEKYSSEKPOYSSEFKKIYEEVYVYDKSILGVSILDSFMYNDLK
PYEKRRVWSTNYCYFWLACFPNINTQIATGLQGLANMQCHITTWIGPGPAG
VVLARVSAAHLSPFISRASFCGIFELWVINKVVAIVYSVOATIAATPVSLML
KSIPEKDLQDKIPDHGSPNATYEFMCFIFMAASLPLFLVPPKHHLFTVAVLV
PPASRGFLIWAIRAHGRIALGSLTDVQPHGSASWAFRLSMGMSFTVINAPV
FSRPEKNSALMSQLVCIPLESLITCLIGLVAAGYEIVGINVSPLDVLEKLOT
TYNKGTRAGVLIISFPAVALQGTNISNSISCTGDMKAIRPKPTINIRGSLFCAMA
ICTCPMNLMATSSKFTMALSAVATPLSSIAGYCSDIYVRRGTYKLTHTIYSHQGSF
TMYGRFGIINRVALAAVLCVAPCLPGTIAVGAPALVDSGAMKLYLSTVWGIGLS
FSSYALCYFPVPVCGPANNIIKQGWFORAWNDPFEKWKDITERDLVDNDISVY
EHEHEKTF1"
6887. .7420
/gene="YBR0304"
6887. .7420

gene
CDS

```

/gene="YBR0304"
/codon_start=1
/protein_id="CA53679.1"
/db_xref="GI:498751"
/db_xref="GOA:P38218"
/db_xref="InterPro:IPR002589"
/db_xref="SGD:S00000226"
/db_xref="UniProt/Swiss-Prot:P38218"
/translation="MSNITYVKNILPKPSVARIILHSQNCNGSWGCIAYQALARP
KAKDVEYCEKXSNLIGKCIILPSEYNSDLICLFTSPGSGSHGKOSILNATK
IALDLKTPRAKDKTRTSEDSIGDYLNHIKYPIGEYELMPQINSGLFGVPMETE
RULBERSGDMSFTVYOL"
/complement(7577..11074)
/gene="CAL1"
/complement(7577..11074)
/gene="CAL1"
/codon_start=1
/product="chitin synthase 3"
/protein_id="CA53680.1"
/db_xref="GI:498752"
/db_xref="GOA:P29465"
/db_xref="InterPro:IPR001199"
/db_xref="InterPro:IPR004835"
/db_xref="SGD:S00000227"
/db_xref="UniProt/Swiss-Prot:P29465"
/translation="MTGLNGDDDDYLYMLNDESLRSHSVSGAPHRQSLVRP
ERSRLNPNPHFYAOKTQOENKHLVLPSTGVNPNATRRSGSLRSKGSVSKFG
RTSDSYLDDMTTDDKASVKISDEGVADDEPDQDVNPEESTOPINKSLPKRK
ERNDTLSFPMQVCYPTTPMAPAPILAFQCMPPKRRPMAREKVALISVILYGAIVAF
LTFGTIVOCSSSKRLKNNVSTEVVINGKAYELDTSSRGIDVEVDSDTLYGPM
SDAGKASFLFQVNGNCHNLITPKSNSSI PHDDNNLAWFPCKLKNODGSKLPFT
VENYAGMNGCHTSKEDDAFYGLKSKADYVFTMGKINSSRNLIYNGVDLIDLDMV
EKDDVYPVFPDLTKTNSQYDLISLVLSNGHBRKARCLSEIIKVEGDSKTVGCI
SDVLYVSLVSLVILIKELIACFRMTVAKQGVAYIDNKMIDKHTIDBMSNII
OKAPLKYVDPHLPRKVKSKSLGHRKASTFOLLKHSGRKQPNESVLDIDMSMS
LOSGSGYRGHTTTTONAKLSNENKAVSRNDSTLLPSSNFKATISPPGSSLI
QSLDSTLIHPDIVQPPDLPMPYGPPLHLITCVTYSDEBELRTLLSTTDIPN
SHLMLAVDCGLIKSGNDKTPPEIAGMMDVTPPEDEKPSYVAVASGSRHNA
KIYAGFYKXDSTIPENQORVPIITVCGTPEAQGAAPGRGRGRDQIILMSFLB
KITPDERMTOLEFOLLKNIWQIGLADAFETVLAWDATKVPDALTVMVAMVDP
LIMGCGETKINKAKSWYTAIOVPEYVISHOAKAPESVFGSVTCLPGCFSMYRIKS
PKSGDGVNVPVLANPDIVERKSDNVNTLHKNLILLGDRPLSSMLTTPPKRVK
VPAACKGTIAPDKFVLLSQRRRINSTVHNLIELVILDLGCTFCFPMQVPIELI
GTWVLPLACFTIYVYIIPAVSKRPTVITLILAIILGPGILVITATRWSTLMMVC
VYICALPIMNFTVPSYAMKFDPDFSMGDTFTIAGNKKAKQDENEGFSDSKIMRTWR
EPEFREDILNRKESDSFVA"
12594..13499
/gene="YBR0308"
12594..13499
/gene="YBR0308"
/codon_start=1
/protein_id="CA53681.1"
/db_xref="GI:498753"
/db_xref="GOA:P38072"
/db_xref="InterPro:IPR003782"
/db_xref="SGD:S00000228"
/db_xref="UniProt/Swiss-Prot:P38072"
/translation="MLNSSRKYACRSLFRANYSIKGLFYNGAVRGRSTGCCASD
NKSPSARQPLDLQGLDEINEPEPIRTFRQPSRKATILALLLSGTYAYISRRR
LLETERKADNARAVGSVALGSPNLDIPFNKSPTEBULKKEBSILFFGFHCDCPE
ELDRLTWVISELDKDHKIQPLFISCDPARDPDVLKELFPHALILGLTETVQV
KSVCKKXVSTPRDYKPNQDYLVDHSIFVYLIDEGQPIDALGVNDYDEOSGLERIR
EJOIAYVPEKERRRSKKWYSFTFN"
/complement(13830..15014)
/gene="YBR0309"
/complement(13830..15014)
/gene="YBR0309"
/codon_start=1
/protein_id="CA53682.1"
/db_xref="GI:498754"
/db_xref="GOA:P38219"
/db_xref="InterPro:IPR004396"
/db_xref="InterPro:IPR006073"

```

```

/db_xref="SGD:S00000229"
/db_xref="UniProt/Swiss-Prot:P38219"
/translation="MPPKQVEKKVLLGRGNNLKAGIVGLANVKGSTFPQAITRCP
LGNPNYPFATIDEPKAVIIPSPFDLCIEIYKKTASVEVPAHLIVDIDLTKGASA
GEGGNAPLISHRSVDSIYQVRCDDMEIIVHGDPVPVADLEIINDELKDIEFA
OKALEGNATIKRGQOSILEVOKKREMLITRKIIKLESSGQVANNHSTSKVEELINS
MFLTAKPCITLINLSESDYIRKXKHLIRIKENWDKXSPDILIIPEVSILBERLSHM
SPEDBEELKQLQTLISALPKITTRRQDLDISFTGCPDEVREMTIRGRTAPADAG
VIHNDLWMTFLIAQWKCEDVEFKYDSDAIKAAGKLMQKRGKQVVEDGDIIFRAGAG
KN"
/complement(16026..17168)
/gene="YBR0310"
/complement(16026..17168)
/gene="YBR0310"
/codon_start=1
/protein_id="CA53683.1"
/db_xref="GI:498755"
/db_xref="GOA:P38071"
/db_xref="SGD:S00000230"
/db_xref="UniProt/Swiss-Prot:P38071"
/translation="MLPTFKRYMSSAQIPEKFKSLIYSTHEVEDCTKVLVKNYTP
KODLSQSIVLKTLAFPIPNSDINLOGVYPERPEKTYDSTDEPAIAGNEGVFVS
LPSGSGKGDILKLDREVIPLAONOGTSMYRVSSSSDLIKVNDLDFSAATVSVNGCT
GEOLVSDYIDNNSNGENWIIIONAGTSVSKITVOYAKKAKITLSVIRDRQFDEYAK
VLEDKYGATKYISQONNDKTPAKEVLSKIIIGENARVALLANSVGKSSASIGARTLEN
NALMTYGGMSQVPTLPTSLIFKGLTSKQVWTERKKKAPQSKITDITDKMTNY
GHIIISPREIEITLWNTNTTDEQLLELVKGIITGKKKQVVLFW"
/complement(17574..19151)
/gene="YBR0312"
/complement(17574..19151)
/gene="YBR0312"
/codon_start=1
/protein_id="CA53684.1"
/db_xref="GI:498756"
/db_xref="GOA:P38070"
/db_xref="InterPro:IPR00719"
/db_xref="InterPro:IPR000961"
/db_xref="InterPro:IPR002290"
/db_xref="InterPro:IPR008271"
/db_xref="InterPro:IPR011009"
/db_xref="SGD:S00000232"
/db_xref="UniProt/Swiss-Prot:P38070"
/translation="WIFSDIEELHVSIDDKNDIKVDSNAIYNDINHEQSSITYE
ESINLVSHNSAIPLNGWSPAIRMRKRSASKEPILPEPNRRSITGSDMRNTN
RSLIPPODIISNIGENELSRNLHDPKRVVGOAYGVLAIVKVTNSKLYAMQAR
KABILISOTATDSKREDEKNDNNNDNDSKLEETFRARSILSEIHEINIYALF
YSFHNLSKVLILQYIPGGLPYHLKHEGTDDETVSPFAARISCALRLHRTKGVYVR"

```

Query Match 97.9%; Score 474.8; DB 15; Length 33117;
 Best Local Similarity 99.6%; Pred. No. 1.8e-80;
 Matches 476; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

| | | | |
|----|------|--|------|
| QY | 6 | CGGAAGGAGGAAGGAGCTCTTCGTCGTGAGACAGTGAATAATTTTACTGTGAT | 65 |
| DB | 7609 | CGGAAGGAGGAAGGAGCTCTTCGTCGTGAGACAGTGAATAATTTTACTGTGAT | 7550 |
| QY | 66 | ACTTACAAGTTGATATATGTTGTGTGTAACCTTATTTTGAAGGATTTTAAACAC | 125 |
| DB | 7549 | ACTTACAAGTTGATATATGTTGTGTGTAACCTTATTTTGAAGGATTTTAAACAC | 7490 |
| QY | 126 | CTTAGAAGCTTAAATTAATTAATTTCTTATCTTTAAAGGACATATTACGTGG | 185 |
| DB | 7489 | CTTAGAAGCTTAAATTAATTAATTTCTTATCTTTAAAGGACATATTACGTGG | 7430 |
| QY | 186 | CTAAGGCAATTCAGCGTATATCTGTAAACATCATGCGCACATTAATCTTCTAACAC | 245 |
| DB | 7429 | CTAAGGCAATTCAGCGTATATCTGTAAACATCATGCGCACATTAATCTTCTAACAC | 7370 |
| QY | 246 | GGCTTCGTCTCTTCCAAAGGAGCTCCGAATATGCCATATTATCTGTGGCAATTC | 305 |
| DB | 7369 | GGCTTCGTCTCTTCCAAAGGAGCTCCGAATATGCCATATTATCTGTGGCAATTC | 7310 |
| QY | 306 | TTTATATTCCTCCTATTGGGATTTGATGCGCGGTTTAAATATGATCACGATTTGATCTTC | 365 |

| ORIGIN | RYLEEFSGDMSFTVYOL" |
|---------------------------------|---|
| Query Match | 92.0%; Score 446.4; DB 15; Length 1089; |
| Best Local Similarity | 99.8%; Pred. No. 1.1e-74; |
| Matches 447; Conservative | 0; Mismatches 1; Indels 0; Gaps 0; |
| Qy | 36 TAGACGATGAAATATTTTCTGCGATCTTACAGAGTTGATATATNGTGTGTGTA 95 |
| Db | 1089 TAGACGATGAAATATTTTCTGCGATCTTACAGAGTTGATATATNGTGTGTGTA 1030 |
| Qy | 96 CTTATTTATTTGAGAGTATTTTAAACAACCTTAGAAGTAACTTAATTAATTAATTT 155 |
| Db | 1029 CTTATTTATTTGAGAGTATTTTAAACAACCTTAGAAGTAACTTAATTAATTAATTT 970 |
| Qy | 156 TCTCTATCTTTAAAGCACTATTACGTGGCTAAGGCAATTACAGCTGATCTGTAAA 215 |
| Db | 969 TCTCTATCTTTAAAGCACTATTACGTGGCTAAGGCAATTACAGCTGATCTGTAAA 910 |
| Qy | 216 ACGTAGTGGCACTAAATCTTCTTACAGGCGTTCGTCTTCCAAAGGAGCTCCGAA 275 |
| Db | 909 ACGTAGTGGCACTAAATCTTCTTACAGGCGTTCGTCTTCCAAAGGAGCTCCGAA 850 |
| Qy | 276 TATGCCACTATTTATCTGTGGCACTTCCAAATTTATTCCTTATTTGGATTTGATGTG 335 |
| Db | 849 TATGCCACTATTTATCTGTGGCACTTCCAAATTTATTCCTTATTTGGATTTGATGTG 790 |
| Qy | 336 GCCGTTTAAATAGTACCGATTTGAATCTTCACTTGTTCGAGTTTGTCTTTGCTTCT 395 |
| Db | 789 GCCGTTTAAATAGTACCGATTTGAATCTTCACTTGTTCGAGTTTGTCTTTGCTTCT 730 |
| Qy | 396 AAGGCTCTCAATTTACTTAAAGCACTTTTGTATTAATTCAAATACTCTTTCCTCC 455 |
| Db | 729 AAGGCTCTCAATTTACTTAAAGCACTTTTGTATTAATTCAAATACTCTTTCCTCC 670 |
| Qy | 456 ATGACTTGAACCTTCCAATGATGAGTAA 483 |
| Db | 669 ATGACTTGAACCTTCCAATGATGATGTA 642 |
| RESULT 4 AY557705/c LOCUS | 534 bp DNA linear PLN 14-MAR-2004 |
| DEFINITION | Saccharomyces cerevisiae clone FLH00707.01X YBR022W gene, complete cds. |
| ACCESSION | AY557705 |
| VERSION | AY557705.1 GI:45269300 |
| KEYWORDS | Yeast ORF Project. |
| SOURCE | Saccharomyces cerevisiae (baker's yeast) |
| ORGANISM | Saccharomyces cerevisiae |
| REFERENCE | Saccharomycetes; Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes. |
| AUTHORS | 1 (bases 1 to 534) Marsischky,G., Rolfs,A., Richardson,A., Kane,M., Bagui,M., Taylor,E., Hu,Y., Vanberg,F., Weger,J., Kramer,J., Moreira,D., Kelley,F., Zuo,D., Raphael,J., Hogle,C., Jepson,D., Williamson,J., Camargo,A., Gonzaga,L., Vasconcelos,A.T., Simpson,A., Kolodner,R., Harlow,E. and LaBaer,J. |
| TITLE | Creation of the YRLEX clone resource: cloning of Saccharomycetes cerevisiae ORFs in the Gateway recombinational cloning system |
| JOURNAL | Unpublished |
| REFERENCE | 2 (bases 1 to 534) |
| AUTHORS | Marsischky,G., Rolfs,A., Richardson,A., Kane,M., Bagui,M., Taylor,E., Hu,Y., Vanberg,F., Weger,J., Kramer,J., Moreira,D., Kelley,F., Zuo,D., Raphael,J., Hogle,C., Jepson,D., Williamson,J., Camargo,A., Gonzaga,L., Vasconcelos,A.T., Simpson,A., Kolodner,R., Harlow,E. and LaBaer,J. |
| TITLE | Direct Substitution |
| JOURNAL | Submitted (17-FEB-2004) Biological Chemistry and Molecular Pharmacology, Harvard Institute of Proteomics, 320 Charles St., Cambridge, MA 02141, USA |
| COMMENT | This clone is part of a collection of Saccharomycetes cerevisiae full-length ORF clones generated by the Harvard Institute of |

| VERSION | KEYWORDS | |
|--|---|--|
| 2358932.1 | GI:536229 | |
| SOURCE | ORGANISM | |
| Saccharomyces cerevisiae (baker's yeast) | Saccharomyces cerevisiae | |
| Eukaryote; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces. | | |
| REFERENCE | AUTHORS | |
| 1 | Feldmann, H., Aigle, M., Ajifiovic, G., Andre, B., Baclet, M.C., Barthé, C., Baur, A., Becam, A.M., Bileau, N., Boles, E., Brandt, T., Brendel, M., Brechner, M., Bressan, F., Christiansen, C., Contreras, R., Crouzet, M., Czepluch, C., Demolis, N., Delaunay, T., Dignon, F., Dondey, H., Duesterhus, S., Dubois, E., Dujon, B., El Bakoury, M., Entian, K.D., Fenerman, M., Fiers, W., Fobo, G.M., Fritz, C., Gassenhuber, H., Glansdorff, N., Goffeau, A., Grivell, L.A., Jacq, C., Heintz, C., Herbert, C.J., Hollenberg, C.P., Holmstrom, K., Jancq, C., Jaquet, M., Jauloux, J.C., Jomiaux, J.L., Kallseer, T., Kleban, P., Kirchbach, L., Koeter, P., Korol, S., Liebl, S., Logghe, M., Lohan, A.D.E., Louis, E.J., Li, Z.Y., Maat, M.J., Maillet, L., Mannheim, G., Messenguy, F., Mlenga, T., Moelens, F., Mueller, S., Naar, F., Obermaier, B., Pereira, J., Pierard, A., Piravandi, E., Pohl, F.M., Pohl, T.M., Potter, S., Proft, M., Purnelle, B., Ramezani, R., Rieger, M., Rose, M., Schaef, Gersenechlaeger, I., Scherens, B., Schwarlose, C., Skala, J., Slonimski, P.P., Smits, P.H.M., Souciet, J.L., Steensma, H.Y., Stucka, R., Vetter, I., Vreeland, A., van der Aart, Q.U., Van Dyck, L., Vasearotti, A., Wolfe, K.H., Zagulski, M., Zimmermann, P.K., Wewes, H.W. and Klein, K. Complete DNA sequence of yeast chromosome II | |
| EMBO J. 13 (24), 5795-5809 (1994) | | |
| JOURNAL | PUBMED | |
| 7813418 | | |
| JOURNAL | REFERENCE | |
| 2 (bases 1 to 5176) | | |
| Grivell, L.A., de Hann, M., Maat, M.J. and Smits, P.H.M. | | |
| JOURNAL | REFERENCE | |
| 3 (bases 1 to 5176) | | |
| JOURNAL | REFERENCE | |
| 1 | MIPS. | |
| Submitted | Direct Submission | |
| Submitted (30-AUG-1994) data collected by MIPS on behalf of the European yeast chromosome II sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152 Martinsried, FRG; E-mail: Mewes@mips.embl.net.org | | |
| FEATURES | source | |
| 1..5176 | /organism="Saccharomyces cerevisiae" | |
| | /mol_type="genomic DNA" | |
| | /db_xref="taxon:4932" | |
| | /chromosome="II" | |
| | complement(160..3657) | |
| | /gene="CHS3" | |
| | complement(160..3657) | |
| | /gene="CHS3" | |
| | /note="ORF YBR023c" | |
| | /codon_start=1 | |
| | /protein_id="CA84965.1" | |
| | /db_xref="GI:536230" | |
| | /db_xref="GOA:P29465" | |
| | /db_xref="SGD:S0000227" | |
| | /db_xref="UniProt/Swiss-Prot:P29465" | |
| | /translation="MTGLNGDDPDYVYLTNLNDDRESILRSHVSGAPHRKOSLVR EKSLRNPPHYAYQKTOEDMNLDTLPSTGVNPAKTRSSLSKSVSKPSG RETDLSLQDMWYTKKASVKISDEGVADDEDKQVDNPESSVTPINKSIPLKAK EINTDLSFQMWYCFITFWAPAPILAFQCPKKERQAMARKVALISVILTAIVAF LTFGPTKVCSSSKLRKNNNEFEVYINGKAYELDPSMFWNKATSSVPSLSL SQAQGDAPFLPQVNGNCHNLITPKSNSIPDDDNNAFYPCPKLNKQDGSKEPNT VNYGAWNCHTSKEDPDAFYGKSAADYFTMDQIGNSRLLIYNGVDVLDLLDD EKDDVDYVYPPDDKTSNIGYDLSLVSNMERKALCSBIIKYNVGVSKYGLCIA SDVILVSLVFLISVLTIKFTIACFKRTVAKQCATYVDKTKDKNTNIDEMSNII OTQAPLEVDPHLRPKKYSKSLGHRKASTPDLKSHSKAFQNESVITDTSWSS LQSGSGRGATTTQTNAMKLSNENKAVHSRNPSTLPTSSMFWNKATSSVPSLSL OLSDSTIHPDIYOQPLDFPMYGPFLHTITFCVYCSGEDEGARTLTDISTDYPN SHGLMWVGDGLIKSGNDKTPBEIALGMDQFVPPDEKVSIVAVASGSKHNA KIYAGYKXDDSTPENNORVPIITIKYCCPGAGAKGNGKRSOIILMSPLE KITDGRMTQLEPOLLKNIWTTGLMADYFVNLVADYTPADLYPDLTHAEMVQD LIMGLCGETKIAKQSWTALQVEFYVISHQAFAESVFGSVTCLPFGCSMRITKS | |

[illegible]

KEYWORDS HTG.
SOURCE Plasmodium falciparum 3D7
ORGANISM Plasmodium falciparum 3D7
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS 1 (bases 1 to 348174)
Hall,N., Pain,A., Berriman,M., Churcher,C., Harris,B., Harris,D.,
Mungall,K., Bowman,S., Atkin,R., Baker,S., Barron,A., Brooks,K.,
Buckee,C.O., Burrows,C., Cherevach,I., Chillingworth,C.,
Chillingworth,T., Christodoulou,Z., Clark,L., Clark,R., Corton,C.,
Crosin,A., Davies,R., Davis,P., Dear,P., Dearden,F., Doggett,J.,
Felwell,T., Goble,A., Goodhead,I., Gilliam,R., Hamlin,N.,
Hane,Z., Harper,D., Hauser,H., Hornsbey,T., Holroyd,S.,
Horrocks,P., Humphray,S., Jagers,K., James,K.D., Johnson,D.,
Kerriourou,A., Knights,A., Kontorov,B., Kyes,S., Larke,N.,
Lawson,D., Lennard,N., Line,A., Maddison,M., McLean,J., Mooney,P.,
Moule,S., Murphy,L., Oliver,K., Ormond,D., Price,C., Quail,M.A.,
Rabbinowitch,E., Rajandream,M.A., Rutter,S., Sutherland,K.M.,
Sanders,M., Simmonds,M., Seeger,K., Sharp,S., Smith,R., Squares,R.,
Squares,S., Stevens,K., Taylor,K., Tivey,A., Unwin,L.,
Whitehead,S., Woodward,J., Sulston,J.E., Craig,A., Newbold,C. and
Barrell,B.G.
TITLE Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13
JOURNAL Nature 419 (6906), 527-531 (2002)
PUBMED 12368867
REFERENCE 2 (bases 1 to 348174)
AUTHORS Cherevach,I., Davis,P., Goodhead,I., Stevens,K., Mungall,K.,
Berriman,M., Pain,A., Hall,N., Atkin,R., Chillingworth,C.,
Doggett,J., Ormond,D., Sanders,M., Hayes,R., Hall,S., Quail,M. and
Barrell,B.G.
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 348174)
Cherevach,I., Davis,P., Goodhead,I., Stevens,K., Mungall,K.,
Berriman,M., Pain,A., Hall,N., Atkin,R., Chillingworth,C.,
Doggett,J., Ormond,D., Sanders,M., Hayes,R., Hall,S., Quail,M. and
Barrell,B.G.
TITLE Direct Submision
JOURNAL Submitted (20-SEP-2002) P.falciparum Genome Sequencing Consortium,
The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK
REFERENCE 4 (bases 1 to 348174)
AUTHORS Cherevach,I., Davis,P., Goodhead,I., Stevens,K., Mungall,K.,
Berriman,M., Pain,A., Hall,N., Atkin,R., Chillingworth,C.,
Chillingworth,C., Doggett,J., Ormond,D., Sanders,M., Hayes,R.,
Hall,S., Quail,M. and Barrell,B.G.
TITLE Direct Submision
JOURNAL Submitted (26-MAR-2004) P.falciparum Genome Sequencing Consortium,
The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK
COMMENT For more information about this sequence or the Malaria Project,
see http://www.sanger.ac.uk/Projects/P_falciparum.
LOCATION/Qualifiers
1. 348174
/organism="Plasmodium falciparum 3D7"
/mol_type="genomic DNA"
/isolate="3D7"
/db_xref="taxon:36329"
/chromosome="6"
complement(847..1143)
/locus_tag="PF0415c"
complement(847..1143)
/locus_tag="PF0415c"
/old_locus_tag="MAL6P1.87"
/codon_start=1
/product="hypothetical protein, conserved"
/protein_id="CAG25326.1"
/db_xref="GI:46361039"
/db_xref="UniProt/TREMBL:Q6LF12"
/translation="MNFSKILPLKXQTIIRNVSLSKNGNHMFKEAATSTENKQ
EOKQVENVYANCCEKGRIRIPRIFGLIFLMAFVPTQSLYETNKYKEVS"
complement(join(2408..3006,3295..3403))
/locus_tag="PF0420c"
complement(join(2408..3006,3295..3403))
/locus_tag="PF0420c"

/old_locus_tag="MAL6P1.88"
/EC_number="3.4.25.1"
/codon_start=1
/product="proteasome subunit alpha type 2, putative"
/protein_id="CAG25327.1"
/db_xref="GI:46361040"
/db_xref="GOA:Q6LFI1"
/db_xref="InterPro:IPR001353"
/db_xref="UniProt/TREMBL:Q6LF11"
/translation="MADGEYSFSLTSPTEKLVQIEYALNRVSSSPALGIRAKGV
IATKSPNLIENSIFPKIQOSEHGIYVAGAPGDFVLLKRAKEAARYSLQYG
SELVLELYTIASTIVQEPOTGVRPGRSLILGVDVGYTHLYQIDPSGCTNMA
TCVGKQVONMSFLEKRNKDIREDIAHTLAITLKESYEGVLANKEIETVAIDNKP
FKILQWEIKYVLEIE"
join(5857..5887,6084..6904)
/locus_tag="PF0425w"
join(5857..5887,6084..6904)
/locus_tag="PF0425w"
/old_locus_tag="MAL6P1.89"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAG25328.1"
/db_xref="GI:46361041"
/db_xref="UniProt/TREMBL:Q6LF10"
/translation="MEDSSLEKSIKGEDESESVNEQNEBSESENESENESENE
NESENEVEENVEENKNEENKNDNMWTOENNNNNNNNNKNDKLNKSFHNNEL
KNLNLYNNINININNNNNNLGYTNLEENGFGQYKNOQDILVNDNKKYSENEKGD
TNHNSKSYKNNENMMIDQKNSKHEKNRNVLPSPSEKRNILFOGNKKEFTP
SLSPDVLIPEKHEHNTMVFHNEPEKGYKKNGLKPSGYLLYPIHSHNSIAPKIK
KKRRCC"
join(9019..10528,10710..10831)
/locus_tag="PF0430w"
join(9019..10528,10710..10831)
/locus_tag="PF0430w"
/old_locus_tag="MAL6P1.90"
/codon_start=1
/product="chaperone, putative"
/protein_id="CAG25329.1"
/db_xref="GI:46361042"
/db_xref="GOA:Q6LFH9"
/db_xref="InterPro:IPR001844"
/db_xref="InterPro:IPR002194"
/db_xref="InterPro:IPR002423"
/db_xref="UniProt/TREMBL:Q6LFH9"
/translation="MSHILNKKADSLSTVNLMTNMAKMYEIIKSNLPGKSYK
MVSSGAIKRTKDGNYLNNEMTIOHPASVLSKSIDETLGGSSNLIYATSLI
VLSRYLILENIHRIITGQFDIKAGLIELDMSKIPVNIENPDKELLVNVCCTCI
RTKPLICLADLDLVESIKIYKPTKQIDLHAIETIDIRANNSINTKLVKGVYLDH
GCRHPNENKLTKEFILLVNLVSLRESVSFVYSNAEDRDLVESERFTDVKV
KTIKIKLIVKEREKTEINIVFAVQKIDPISLDLAKENIMALARIKRRLNERI
VLCGAPCNVVDLTEDVDVAGLYVEISINDKYTEIEVQNPSCSTIOANDY
TIIKQIKDRIQGLSKIVNIDKCVLSAGSFEIYMAKYKLDKEKKIKGKQFALDIY
ANSLNIPKVLINSGLDIHOTLFNVIDKNEDSEPLGLDLTGEPIIHLKGIYDN
YCVYKEIISTATIASQILLVDIIRAKSGEER"
13263..14507
/locus_tag="PF0435w"
13263..14507
/locus_tag="PF0435w"
/old_locus_tag="PF0435w"
/old_locus_tag="MAL6P1.91"
/EC_number="2.6.1.13"
/codon_start=1
/product="ornithine aminotransferase"
/protein_id="CAG25330.1"
/db_xref="GI:46361043"
/db_xref="GOA:Q6LFH8"
/db_xref="InterPro:IPR005814"
/db_xref="InterPro:IPR010164"
/db_xref="UniProt/TREMBL:Q6LFH8"
/translation="MDFPYELKSSQDVNNNMLTYGAHNYDPIPVYLKRGKGVFYVDIE
DRRYVDFLFASSVNOGHCHDILNAMIQAKLTI CSRAPFSLSIGYCEBYLNLPG
YKVLVMTAGASSETAYKLCRKNGYERKIPENSAKTIYCNPNNSGSLGCVASSTV
KKCKNRGPPVYPLKVPYDDELALAELEODPNVCAKIVVEVQEGAGYIVSDSYFPG
VASLCRTYVNLVADDEVDTGIGRTGKLCTHHYGVKPDVILIGKALSGGHPISAILLA

NNDWMLVKRHHGSGTYGNNPLAAATCVAELKYLINRKLCEBNADKGAFLONLKEOL
 KQSVREVRAGKGLCLIEFKNDLVNWMIDCLKFKXNGLITRSHKTRALRPLPLIT
 KEQLDCTETLIVTKFEPDNL"
 /join(15915..18857,19043..19087)
 /locus_tag="PF0440w"
 /join(15915..18857,19043..19087)
 /locus_tag="PF0440w"
 /old_locus_tag="MAL6P1.92"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="CAG25331.1"
 /db_xref="GI:46361044"
 /db_xref="UniProt/TREMBL:O6LPH7"
 /translation="MNYGICDYLKKKSSSYLLETNRYKKLMNNEGDNPNFKIK
 INLTKENESTIKMKRLILKRPDIPTVSTYSRLKKNIEENHIIIPDICKCYEMKG
 IFESDEFEVYVELKLYCDKNQVNNIENNVDNNIDDKYIDDKYIDDKYIDN
 KYVDNKKYVDGMLGFYDNNNPIDDDFINNNISPLNI PVLYKRNQNNKGCIL
 YIKMSIYNNEFLANKYTYINYKDWGLKYLKQYKRMVAKQKTEINIVNMSDYL
 ETLKEVLLYCVVVDYKRYKKNMDCIKVLNNMKDYKSMVNDIYNFKNMEYV
 NEYKATPLMKHNIILKSYKEPIKNTSHINDYVYDILNFPQOINKLKNMEYPI
 FTGMDNLYKHIIITVSVCLINIKTNPFTHNQLSAVKKDYLITLKNMEYILN
 NKNDNHNNNYINNNYINDVVDYQDNIHMTSPSYGNSVDVDFPLNTCNAQTK
 DEKVEENNNNIKNIMLNMMNKKKTCEYFDLHPNIIKPLNINSCIFNNMYH
 VILKSKDINGIDFNDIYKDEETVDNNKIYANNYNNDPSDHYNKEEHNYSKKT
 KNOKKKTSYLDYINRKTMMGTYDYEDKMLKNHNFIDILYEERKKKDKRNNI
 QIKSTNNQNMHYSCKTNESEYTHNKPININSNEIKKCNVEYISIDNNIRBY
 KIKESILNLFYRKVEKNSYESKTYEQDVSIIFPNILAKFLVYIINNPLQSNPF
 PPOVWLLFNSIDOKIKRQTTNNNTKNIQVLEHILKDEKNRILHYTHDPSLITF
 NVNOEKENIKRMTDICKDNIMYEKISTYEHPLIYFENHLLFLNKLKISYDLS
 SYVINLPENYSKNLIRTIILITFNQVFTSPILQYLQFELLIFPFSYFVLSI
 PKADGSTERIKFVLSYGSLS"
 20889..39122
 /locus_tag="PF0445w"
 20889..39122
 /locus_tag="PF0445w"
 /old_locus_tag="MAL6P1.93"
 /note="Asparagine-rich"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="CAG25332.1"
 /db_xref="GI:46361045"
 /db_xref="InterPro:IPR001444"
 /db_xref="InterPro:IPR006162"
 /db_xref="UniProt/TREMBL:O6LPH6"
 /translation="MKNDRGKKGEPRTFYETIKKLFVYECALINENRKNRKLK
 KKKKIESLKLQDILLIFDNNIDILCDPESFLYEEEDVAKTEIKATYISNYENL
 NMDIILKCSILINRAVYVCTLFNDVYLKIKESNIKNEIMKCTYICNSDFEKLEN
 LTIMYELIKYINKRERIGVSLNNDKQYKYEETINSTDKDNNSSDDNNGIILK
 NEKVTSSVHSTTNNISVGSINISKTVEENINNNRVNSDRFSYIILILQNDPID
 NIFLEIRIKFLIKYDILCDQINIVFVNTKHKINILFLDLILFLYSKFOATYQII

Query Match 10.8%; Score 52.2; DB 2; Length 348174;
 Best Local Similarity 57.8%; Pred. No. 0.63;
 Matches 93; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 39 ACAGTATGAAATATTTTACTGTATCTTACAAGTTGATATATGCTGTGTGTAAGT 98
 DB 74779 ACATTAAGAAAAATTTATACATGTTATATACATATATATATATATATATAT 74838
 QY 99 ATTATATTGAGAGTATTTTAACACACCTTAGAATTAAGTAAATTAATATTTTCT 158
 DB 74839 TATTATTTTATAGTATTTTAAAAAAGTCTATATATTTATATATAT 74898
 QY 159 CTATCTTAAAGCACATATTACGTGGCTAAGCAATTACA 199
 DB 74899 ATATATATATATATATACAGCTAATATACAGAAATAATAA 74939

RESULT 9
 BV061001 633 bp DNA linear STS 31-MAY-2003
 LOCUS S212P6318FG5.T0 CZECHII/BI Mus musculus STS genomic, sequence
 DEFINITION tagged site.

ACCESSION BV061001 GI:31176796
 VERSION BV061001.1
 KEYWORDS STS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Muridae; Murinae; Mus.
 1 (bases 1 to 633)
 Waded, C.M., Kulbokas, E.J. III, Kirby, A.W., Zody, M.C., Mullikin, J.C.,
 Lander, E.S., Lindblad-Toh, K. and Daly, M.J.
 The mosaic structure of variation in the laboratory mouse genome
 Nature 420 (6915), 574-578 (2002)
 12466852

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 PUBMED
 COMMENT

Contact: Kerstin Lindblad-Toh
 Whitehead Institute for Biomedical Research, Center for Genome
 Research
 320 Charles Street, Cambridge, MA 02141, USA
 Tel: 6172521477
 Fax: 6172580903
 Email: kersti@genome.wi.mit.edu
 Primer A: None
 Primer B: None
 STS size: 633
 Protocol:
 WGS-discovery: Paired-end low-coverage whole genome shotgun reads
 were generated from 129S1/SvJm, C3H/HeJ, and BALB/cByJ. The WGS
 reads were placed uniquely on the MGSCv3 C57BL/6J assembly and SNP
 detection was carried out by SNAHA-SNP. 225,000 reads were
 annotated
 as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J
 and the strain from which the particular read came. The validation
 rate for these SNPs was estimated at approximately 98%.
 Location/Qualifiers
 1..633
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="CZECHII/BI"
 /db_xref="taxon:10090"
 /map="5-33-576 61053125-61052582"
 /clone_lib="CZECHII/BI"
 <1..>633

ORIGIN
 STS

Query Match 10.6%; Score 51.4; DB 10; Length 633;
 Best Local Similarity 48.2%; Pred. No. 5.7;
 Matches 145; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY 17 AAGTCACTCTTCGTTGGGTGAGACAGTATGAAATATTTTACTGTGATCTTACAAGTT 76
 DB 407 AAAACACATTTTATTTTATATATACATTTATTAACATGACTTACTGTGACATTTAT 348
 QY 77 GATATATGTTGTGTGTAACCTTATTTATTTAGAGGTATTTTACACACCTTAGAAGTAA 136
 DB 347 CATATGCAATCATGAGATCTTAATTAATGTAATTTTATATACAGCACCTTAACAGTCA 288
 QY 137 AACTTAATTAATTAATTTTCTATCTTTAAAGCACATATTAAGTGGCTAAGCAATT 196
 DB 287 CAGTTATATTTTAAATCTTGTAATTTTGTATATGTAATAATTAATTTTAACTACTTTT 228
 QY 197 ACAGCTGATATCTGTAAAACTCATGTGCGCACTAAATCTTCTTAACACGCGTGTGCT 256
 DB 227 TAAATTAATTTTTCGAAATTTTCGGAGATGTGATGTGATTCACATGACTTTTACCCCT 168
 QY 257 CTTTCCAAAGGACCTCGAATATGCACTATTTATCTGTGGCACTTTCGCAATTTATATGCC 316
 DB 167 CCTTCAATTTTACCCCTTCCTCAACCTCTTCAACCTTTCATTCACAGGTTTCATGACA 108
 QY 317 C 317
 DB 107 C 107

RESULT 10
AC079954 158785 bp DNA linear PRI 19-MAR-2003
LOCUS Homo sapiens 12 BAC RPL1-179A1 (Roswell Park Cancer Institute Human
DEFINITION BAC library) complete sequence.
AC079954
VERSION AC079954.18 GI:13899365
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukayocsa, Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
1 (bases 1 to 158785)
REFERENCE
AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alshbrook, S.L., Amaratunga, H.C., Are, J.R., Banks, T., Barbarella, J.,
Benton, J., Binage, K., Blankenburg, K., Bonnin, D., Bouck, J.,
Bowle, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Bunay, C.,
Burich, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chiu, D., Chowdhry, I., Christopoulos, C., Cleveland, C.D.,
Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Day-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Emerling, S., Escotto, M., Falls, T., Ferraguto, D.,
Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J.,
Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W.,
Gunnarsson, P., Hale, S., Hamilton, K., Han, J., Harris, C., Harris, K.,
Hart, M., Havlak, P., Hawes, A., Hernandez, U., Hernandez, O.,
Hodgson, A., Hogue, M., Hollway, C., Hollins, B., Homsl, F.,
Howard, S., Huber, J., Huiyk, S., Hume, J., Ioshikhes, I., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlsone, E., Kelly, S., Khan, U., King, L., Korvach, J., Kovar, C.,
Kratovic, C., Kureshi, A., Landry, N., Leal, B., Lee, E., Lewis, L.C.,
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lien, C., Liu, J., Liu, W.,
Loudeseg, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Marondel, I., Martin, R.,
Martindale, A., Martinez, B., Massey, E., Mawhinney, E., McLeod, M.P.,
Meador, M., Mei, G., Merscher, S., Metzger, M., Miller, A., Miner, G.,
Miner, Z., Mitchell, T., Mohabbat, K., Montgomery, K.T., Morgan, M.,
Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N.,
Nguyen, S., Nguyen, N., Nguyen, N., Nickerson, E., Nwokweto, S.,
Ogih, M., Okunou, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,
Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shin, C.,
Shooshitari, N., Sisson, I., Sodergren, E., Sonalike, T., Sparks, A.,
Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A.,
Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B.,
Thomas, N., Thomas, S., Umanan, K., Vasquez, L., Vera, V., Villalon, D.,
Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R.,
Washington, C., Watlington, S., Williams, G., Williamson, A.,
Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J.,
Zorrilla, S., Zuchterlapati, R. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 158785)
REFERENCE
AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (20-SEP-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 158785)
REFERENCE
AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (01-MAY-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 158785)
REFERENCE
AUTHORS Worley, K.C.
TITLE Direct Submission

JOURNAL
Submitted (07-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 158785)
REFERENCE
AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (25-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
6 (bases 1 to 158785)
REFERENCE
AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (19-MAR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 1, 2001 this sequence version replaced gi:13877202.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:
STS are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as low coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards. Estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL:
<http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

| Summary Statistics | |
|--|-------------|
| Contig length: | 158785 |
| Phrap values in estimate: | 158572 |
| Average error rate (BCM-Phrap estimate): | 4.46022e-06 |
| Fraction of Phrap values less than 40 : | 0.00425044 |
| Number of consensus changing edits: | 30 |
| Number of N's in consensus : | 0 |

----- Consensus changing edits -----

| Position | Original+Context | Edited+Context |
|----------|-------------------------|------------------------|
| 10464 | ctctctctc(n)ctnctctctc | ctctctctc(t)ctctctctc |
| 10467 | ctctctctc(n)ctnctctctc | ctctctctc(c)ctctctctc |
| 10469 | ctctctctc(n)ctnctctctc | ctctctctc(c)ctctctctc |
| 10491 | ctctctctc(n)ctnctctctc | ctctctctc(c)ctctctctc |
| 10511 | ctctctctc(n)ctnctctctc | ctctctctc(c)ctctctctc |
| 13930 | taccaccatc(n)ntnnnncaag | taccaccatc(c)ctgcacaag |

| | | |
|--------|---------------------------|---------------------------|
| 13931 | accacactm(n) tmmnaagaac | accacacttc(c) tgaacaaga |
| 13933 | ccacactcm(n) mmnaagaac | ccacactccc(g) caacaagaac |
| 13934 | ccactcmtn(n) mnaagaac | ccactccctc(t) caaagaac |
| 13935 | cactcmtn(n) ncaagaact | ccactccctgc(a) acaagaact |
| 13936 | actcmtnm(n) caagaacct | actccctgca(a) caagaacct |
| 37060 | aaccaaatc(m) gyscaatagt | aaccaaatc(t) tgcgaatagc |
| 37063 | attnggcaa(n) tgaatccaac | atttggcaa(l) tgaatccaac |
| 37085 | aacaatatat(n) caagatbatc | aacaatatat(a) gaagatbatc |
| 37109 | tatbactraa(n) gaattttacc | tatbactraa(l) gaattttacc |
| 41956 | gyacacataa(n) caggtcttca | gyacacataa(g) caggtcttca |
| 51989 | aagyccctca(n) agaaagaytc | aagyccctca(g) aaagaaaytc |
| 63302 | atcccatctc(n) aaaaaaagtc | atcccatctc(a) aaaaaaagtc |
| 64961 | gtatctctag(t) tcttttttt | gtatctctag(t) tcttttttt |
| 64967 | ctaantctc(n) ttttttttt | ctaantctc(t) ttttttttt |
| 65267 | actgcaccca(n) ccacatttta | actgcaccca(g) ccacatttta |
| 75281 | cttatrttgg(t) aacttgltta | cttatrttgg(l) aacttgltta |
| 92134 | aagrgagggg(n) gccrttttaa | aagrgagggg(a) gccrttttaa |
| 102189 | cttgttgcuaa(n) gytttcagtg | cttgttgcuaa(g) gytttcagtg |
| 117252 | aacccacatgt(n) lpgaacsctg | aacccacatgt(g) lpgaacsctg |
| 117431 | aacagcgaaq(n) naacaagaat | aacagcgaaq(a) gaacaagaat |
| 117432 | aacagcaagan(n) aacaagaact | aacagcaagan(g) aacaagaact |
| 144804 | tgsagtltag(n) gyscaacttc | tgsagtltag(t) gyscaacttc |
| 144939 | gtatrtttag(t) agagaagag | gtatrtttag(t) agagaagag |
| 158784 | ctttcaaggaa(a)t | ctttcaaggaa(t)t |

| Distribution of Quality < 40 Bases | |
|------------------------------------|---|
| 500 | * |
| 450 | * |
| 400 | * |
| 350 | * |
| 300 | * |
| 250 | * |
| 200 | * |
| 150 | * |
| 100 | * |
| 50 | * |
| 0 | * |

```
-----
Version: 1.01  gxf0.
```

```
misc_feature      1., 1977
                  /note="overlaps bases 105124. .107100 of clone AC113777"
misc_feature      /function="clone overlap"
                  1375., .1611
```

| | | | | |
|-----------------------|-----------------|-----------------|-----------|----------------|
| Query Match | 10.6%; | Score 51.4; | DB 8; | Length 158785; |
| Best Local Similarity | 49.6%; | Pred. No. 1.1; | | |
| Matches 186; | Conservative 0; | Mismatches 186; | Indels 3; | Gaps 2; |

| QY | Db | QY | Db | QY | Db |
|---|--|--|---|--|---|
| 43 | 135259 | 103 | 135319 | 163 | 135379 |
| ATGAAAATATTTTACTGTGATCTTACAAAGTGAATATAGGTGTGTACTAAATT | TATGACTTTTCTTATCCTTGAAATTATAAAAATTAAAAATCCCTCTAAATATATTT | ATTGAGAGTATTTTAACACACTAGAACTAAACTTAAATAAATATTTCTCTAT | ATTTCAGAAATTTTCTAAATTCAGTAAATCTGTAATTTATTTCTAGGCATTTCTTAA | CTTTAAAGCACAATTAACGTGGCTTAAGCAATTAACGCGATATACGTGAAAACCTATG | C-TCATTAACCAACATGAGAGTATTTCTTTATTTGATGCTTTATCTTCACTGCTACACTTT |
| 102 | 135318 | 162 | 135378 | 222 | 135437 |

| | | | |
|----|--------|--|--------|
| OY | 223 | TGCGCAGCAAAATTCCTTCCAAACGGGTCTGCCTCTTCCAAAGGAGCCGAATTAGCCA | 282 |
| Db | 135438 | TCGTTCTAAAATTCCTATCATATGACATTTGGATC--TCATGGAATCTTTCCCCTGTTT | 135495 |
| OY | 283 | CTATTATCTGTGGCATTTCCAAITTTATATGCCCTATGGGTAATTTGATGAGCCGTTT | 342 |
| Db | 135496 | TATCTTTCTCTTGAAATTTTTATTTATAACTTCTCTGCTAATTAATGCTTAAATAT | 135555 |
| OY | 343 | AAATATGTCACCGATTGAATCTTCACTTGTTCAGATTTTGTCTTTGCTCTCTCAAAGATC | 402 |
| Db | 135556 | TCTACTTACGCACTGGGGTTTCTCATATCCAGTATTTTGTGCTTCTTTACAGTTT | 135615 |
| OY | 403 | TTCAATTTATCTAAA | 417 |
| Db | 135616 | TGAATTTGATATGAA | 135630 |

| | | | | | |
|------------|---|-------------|-----|--------|-----------------|
| RESULT | 11 | | | | |
| LOCUS | AC123054/c | | | | |
| DEFINITION | Mus musculus BAC clone RP24-220N11 from 5, complete sequence. | 174874 bp | DNA | linear | ROD 05-NOV-2003 |
| ACCESSION | AC123054 | | | | |
| VERSION | AC123054.2 | GI:22213530 | | | |
| KEYWORDS | HTG. | | | | |
| SOURCE | Mus musculus (house mouse) | | | | |
| ORGANISM | Mus musculus | | | | |

| | |
|-----------|---|
| REFERENCE | 1 (bases 1 to 174874) |
| AUTHORS | Nguyen,C., Bielicki,L. and Creason,K. |
| TITLE | The sequence of Mus musculus BAC clone RP24-220N11 |
| JOURNAL | Unpublished (2001) |
| REFERENCE | 2 (bases 1 to 174874) |
| AUTHORS | Wilson,R. |
| TITLE | Sequencing of Mus musculus |
| JOURNAL | Unpublished (2001) |
| REFERENCE | 3 (bases 1 to 174874) |
| AUTHORS | McPherson,J.D. and Waterson,R.H. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (27-MAY-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA |
| REFERENCE | 4 (bases 1 to 174874) |
| AUTHORS | McPherson,J.D. and Waterson,R.H. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA |
| REFERENCE | 5 (bases 1 to 174874) |
| AUTHORS | McPherson,J.D. and Waterson,R.H. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (13-AUG-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA |
| REFERENCE | 6 (bases 1 to 174874) |
| AUTHORS | Wilson,R. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (05-NOV-2003) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA |
| COMMENT | On Aug 13, 2002 this sequence version replaced gi:121217626. |

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: submissions@wustl.wustl.edu
----- Summary Statistics
Center project name: M_IB0220N11

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCT-24 BAC Library has been constructed by Pieter de Jong and coworkers (<http://www.chori.org>) from male C57Bl/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

FEATURES
Location/Qualifiers

```
1..174874
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="5"
/map="5"
/clone="RP24-220N11"
/clone_1b="RPCT-24"
1..481
/rpt_family="L1"
repeat_region
1533..1639
/rpt_family="L1"
repeat_region
1635..2294
/rpt_family="L1"
repeat_region
2312..2541
/rpt_family="L1"
repeat_region
2551..3982
/rpt_family="L1"
repeat_region
3980..5313
/rpt_family="L1"
repeat_region
5404..7065
/rpt_family="L1"
repeat_region
7314..8273
/rpt_family="L1"
repeat_region
8831..8883
/rpt_family="ERV1"
repeat_region
9597..9729
/rpt_family="L1"
repeat_region
10923..11059
/rpt_family="Alu"
repeat_region
11394..11592
/rpt_family="MER1_type"
repeat_region
11914..12004
/rpt_family="L1"
repeat_region
12002..13641
/rpt_family="L1"
repeat_region
13517..13573
/rpt_family="ERV1"
repeat_region
15824..15915
/rpt_family="Alu"
repeat_region
16175..16378
/rpt_family="L1"
repeat_region
16379..16526
/rpt_family="Alu"
repeat_region
16527..16808
/rpt_family="L1"
repeat_region
16805..17645
/rpt_family="L1"
repeat_region
17893..17982
/rpt_family="MER2_type"
repeat_region
18413..18629
/rpt_family="L1"
repeat_region
18689..19125
/rpt_family="L1"
repeat_region
19126..19511
/rpt_family="MALR"
repeat_region
19512..19532
/rpt_family="L1"
repeat_region
24546..24780
/rpt_family="L1"
repeat_region
24817..25216
/rpt_family="L1"
repeat_region
25285..25831
/rpt_family="L1"
repeat_region
27096..27471
/rpt_family="MALR"
repeat_region
28325..28446
/rpt_family="B4"
repeat_region
28357..28679
/rpt_family="MALR"
repeat_region
28844..29149
/rpt_family="MALR"
repeat_region
29164..29377
/rpt_family="L1"
repeat_region
30082..30643
/rpt_family="L1"
repeat_region
30806..31168
/rpt_family="MALR"
repeat_region
31398..31799
/rpt_family="L1"
repeat_region
34685..34821
/rpt_family="MALR"
repeat_region
34714..34835
/rpt_family="Alu"
repeat_region
34871..35975
/rpt_family="L1"
repeat_region
35981..36481
/rpt_family="L1"
repeat_region
36468..36857
/rpt_family="L1"
repeat_region
36863..37796
/rpt_family="L1"
repeat_region
37826..38542
/rpt_family="L1"
repeat_region
38875..38977
/rpt_family="L1"
repeat_region
38992..39372
/rpt_family="L1"
repeat_region
39377..39564
/rpt_family="ERVK"
repeat_region
39466..39660
/rpt_family="B4"
repeat_region
39664..39985
/rpt_family="L1"
repeat_region
41371..41452
/rpt_family="L1"
repeat_region
41714..41854
/rpt_family="Alu"
repeat_region
42684..42717
/rpt_family="L1"
repeat_region
42713..43611
/rpt_family="L1"
repeat_region
44642..45115
/rpt_family="L1"
repeat_region
45116..45922
/rpt_family="L1"
repeat_region
46124..46424
/rpt_family="L1"
repeat_region
47263..47408
/rpt_family="Alu"
repeat_region
48705..49182
/rpt_family="L1"
repeat_region
49585..49839
```

| | | |
|---|---|--------------------------------------|
| | /ipc_family="L1" | |
| Query Match | 10.6%; Score 51.4; DB 9; Length 174874; | |
| Best Local Similarity | 48.2%; Pred. No. 1.1; | |
| Matches 145; Conservative | 0; Mismatches 156; Indels 0; Gaps 0; | |
| OY | 17 AAGTACCCCTCGTGGTGATGAAGAATAATTACTGGTAGTAACAAGTT | 76 |
| Dd | 22376 AAAAACAATTTATTTTTATTTATTAATTTAACTCAATTCAGCATTACCATTT | 22311 |
| OY | 77 GATATATGGTGTGTGTAACTATTATTGAGAGTATTTAACACCTTGAACATA | 136 |
| Dd | 22316 CATATGCATATCATAGAGATCTAATTAATATGTAATTTTCAATATGCCACTAAACATCA | 2225 |
| OY | 137 AACCTAATAAATAATATTTCTCTATCTTTAAGGCACATTTACGTGGCTAAGGCAATT | 196 |
| Dd | 22256 CAGTTATATTTTAAATCTCTGTAAATTTAGTATATGAAAAATTAATTTTACTCTT | 22197 |
| OY | 197 ACAGCTGATATACGTAAATCTCATATGTCGCACAAAATTCTGTCAAACGCGTTGCTG | 256 |
| Dd | 22196 TAAATTAATATTTCTGAAAATTTCTGGAGTGTATGTATTCACATCACCTTTACCCT | 2213 |
| OY | 257 CTTCGCAAGGACCTCCGAATANGCAGCATATTATCTGTGGCATTTCCAATTTAATCCC | 316 |
| Dd | 22136 CCTTCTCAATTTTACCCCTCCTTCAACCTCTTCCATTGGCATTACAGAGTTCCATCACA | 22077 |
| OY | 317 C 317 | |
| Dd | 22076 C 22076 | |
| RESULT 12 | | |
| LOCUS | AC124497 | 206926 bp DNA linear ROD 11-NOV-2003 |
| DEFINITION | Mus musculus BAC clone RP23-480E8 from chromosome 5, complete sequence. | |
| ACCESSION | AC124497 | |
| VERSION | AC124497.3 GI:28016157 | |
| KEYWORDS | HTG. | |
| SOURCE | Mus musculus (house mouse) | |
| ORGANISM | Mus musculus | |
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Mus. | | |
| REFERENCE AUTHORS | Swearingen-Shahid,S., Cotton,M., Shahid,S., Meyer,R. and Bielik,L. The sequence of Mus musculus BAC clone RP23-480E8 Unpublished (2001) 2 (bases 1 to 206926) Wilson,R. Sequencing of Mus musculus Unpublished (2001) 3 (bases 1 to 206926) McPherson,J.D. and Waterston,R.H. Direct Submission Submitted (14-JUN-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA 4 (bases 1 to 206926) McPherson,J.D. and Waterston,R.H. Direct Submission Submitted (26-NOV-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA 5 (bases 1 to 206926) McPherson,J.D. and Waterston,R.H. Direct Submission Submitted (29-JAN-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA 6 (bases 1 to 206926) Wilson,R. Direct Submission Submitted (11-NOV-2003) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA | |

```

COMMENT
----- Genome Center
On Jan 29, 2003 this sequence version replaced gt:25450603.

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watsn.wustl.edu
----- Summary Statistics
Center project name: M_BA0480E08
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Mes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RRC1-23 BAC library has been constructed by Kazuroyo Oseegawa
and Minako Tatemio in the laboratory of Pieter de Jong
(http://www.chori.org) from female C57BL/60 mouse kidney and/or
brain genomic DNA. The clone and detailed information can be
obtained from Research Genetics, Inc. (http://www.resgen.com) or
Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone. This clone is
overlapped by AC123054.

FEATURES
source          1..206926
                location/Qualifiers
                /organism="Mus musculus"
                /mol_type="Genomic DNA"
                /db_xref="taxon:10090"
                /chromosome="5"
                /map="5"
                /clone="RP23-480B8"
                /clone_lib="RRC1-23"
                /size..1952
repeat_region   /rpt_family="r1"
                2182..2544
repeat_region   /rpt_family="MaLR"
                2826..3268
repeat_region   /rpt_family="r1"
                3973..4186
repeat_region   /rpt_family="r1"
                4201..4505
repeat_region   /rpt_family="MaLR"
                4670..4792
repeat_region   /rpt_family="MaLR"
                4903..5027
repeat_region   /rpt_family="B4"
                5878..6253
repeat_region   /rpt_family="MaLR"
                7516..8054
repeat_region   /rpt_family="r1"
                8133..8532
repeat_region   /rpt_family="r1"
                8569..8803
repeat_region   /rpt_family="r1"
                13817..13837
repeat_region   /rpt_family="r1"

```

| | | |
|---------------|--------------|--------------------------|
| repeat_region | 13838..14223 | /rpt_family="MaLR" |
| repeat_region | 14234..14660 | /rpt_family="L1" |
| repeat_region | 14720..14996 | /rpt_family="L1" |
| repeat_region | 15367..15436 | /rpt_family="MaLR2_type" |
| repeat_region | 15704..16544 | /rpt_family="L1" |
| repeat_region | 16541..16882 | /rpt_family="L1" |
| repeat_region | 16823..16970 | /rpt_family="Alu" |
| repeat_region | 16971..17114 | /rpt_family="L1" |
| repeat_region | 17434..17555 | /rpt_family="Alu" |
| repeat_region | 17776..17832 | /rpt_family="ERV1" |
| repeat_region | 19708..21347 | /rpt_family="L1" |
| repeat_region | 21345..21435 | /rpt_family="L1" |
| repeat_region | 21577..21955 | /rpt_family="MER1_type" |
| repeat_region | 22290..22466 | /rpt_family="Alu" |
| repeat_region | 23620..23752 | /rpt_family="L1" |
| repeat_region | 24455..24518 | /rpt_family="ERV1" |
| repeat_region | 25076..26035 | /rpt_family="L1" |
| repeat_region | 26284..27945 | /rpt_family="L1" |
| repeat_region | 28036..29369 | /rpt_family="L1" |
| repeat_region | 29367..30798 | /rpt_family="L1" |
| repeat_region | 30808..31037 | /rpt_family="L1" |
| repeat_region | 31055..31714 | /rpt_family="L1" |
| repeat_region | 31710..31816 | /rpt_family="L1" |
| repeat_region | 32868..34208 | /rpt_family="L1" |
| repeat_region | 34267..34341 | /rpt_family="MaLR" |
| repeat_region | 35856..35925 | /rpt_family="L1" |
| repeat_region | 36971..37274 | /rpt_family="MaLR" |
| repeat_region | 37824..37906 | /rpt_family="ERV1" |
| repeat_region | 38023..38355 | /rpt_family="ERV1" |
| repeat_region | 38655..39782 | /rpt_family="ERV1" |
| repeat_region | 39782..39923 | /rpt_family="L1" |
| repeat_region | 39924..40316 | /rpt_family="MaLR" |
| repeat_region | 40317..41412 | /rpt_family="MaLR" |
| repeat_region | 41413..41806 | /rpt_family="MaLR" |
| repeat_region | 41807..42069 | /rpt_family="L1" |
| repeat_region | 44555..44675 | /rpt_family="L1" |
| repeat_region | 48259..48354 | /rpt_family="L1" |

| Query Match | 10.6% | Score 51.4 | DB 9 | Length 206926 |
|-----------------------|---|--|----------|-----------------|
| Best Local Similarity | 48.2% | Fold. No. 1 | | |
| Matches 145 | Conservative 0 | Mismatches 156 | Indels 0 | Gaps 0 |
| repeat_region | /rpt_family="L1" | 48777..49054 | | |
| repeat_region | /rpt_family="ERV.L" | 49190..50095 | | |
| repeat_region | /rpt_family="L1" | 50111..50185 | | |
| repeat_region | /rpt_family="L1" | 50178..50811 | | |
| repeat_region | /rpt_family="L1" | 50812..51133 | | |
| repeat_region | /rpt_family="L1" | 51765..51952 | | |
| repeat_region | /rpt_family="B4" | 52274..52574 | | |
| Db | 17 | AAGTACTCCTTGCTGGTGCAGTACAGTAAATATTTTACTGTGATCTTACAGTT | 76 | |
| Db | 10973 | AAAACATTTTATTTTATATATCATATTTTACATGACTTCTGACATTTACATATT | 110322 | |
| Qy | 77 | GATATANGTGTGTGTAATCTTATTTATTTGAGAGTATTTTAAACACCTTAGACTTA | 136 | |
| Db | 11033 | CATATGCATATCATGGAATCTATATATATTTTATATATGATTTTATATGACGACCTTAAACATCTA | 110922 | |
| Qy | 137 | AACCTAATTAATTAATTAATTTCTCTATCTTTTAAAGCACATATTTACGTGCTAAGCAATT | 196 | |
| Db | 11093 | CAGTTATATTTTATATTTCTTGTATTTTATGATATATGAAATTAATTAATTTTACACTTTT | 111522 | |
| Qy | 197 | ACAGCTGATATCTGTAAACATCATGTGCGCACTAATTTCTTCTAAGCGCTTCTGTCT | 256 | |
| Db | 11153 | TAAATTAATTTTCTGAAATTTCTGCAATGTGTATTTGATTCACATCTTTTACCCTT | 11212 | |
| Qy | 257 | CTTTCGAAGGACCTCGAATATGCGCATTTATCTGTGCGCATTTCAATTAATATTCCTC | 316 | |
| Db | 11213 | CTTTCGAATTTTACCTCTCTTCAACCTTCTTCATTTGCGATTTACAGGTTTCATTCACA | 11272 | |
| Qy | 317 | C 317 | | |
| Db | 11273 | C 11273 | | |
| RESULT 13 | | | | |
| AC087277 | 196859 bp | DNA | linear | PRI 24-APR-2002 |
| LOCUS | AC087277 | | | |
| DEFINITION | Homo sapiens chromosome 11, clone RP11-514F3, complete sequence. | | | |
| ACCESSION | AC087277 | | | |
| VERSION | AC087277.11 | | | |
| KEYWORDS | HTG. | | | |
| SOURCE | Homo sapiens (human) | | | |
| ORGANISM | Homo sapiens | | | |
| REFERENCE | Bukharin, V., Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi, Mammalia, Eutheria, Euarchontoglires, Primates, Catarrhini, Hominidae, Homo. | | | |
| AUTHORS | 1 (bases 1 to 196859) | | | |
| TITLE | Homo sapiens chromosome 11, clone RP11-514F3 | | | |
| JOURNAL | Unpublished | | | |
| REFERENCE | 2 (bases 1 to 196859) | | | |
| AUTHORS | Birken, B., Linton, L., Nubaum, C., Lander, E., Allen, N., Anderson, S., Batta, N., Beattie, V., Boguslavsky, L., Bouhassira, B., Brown, A., Camarata, J., Campopiano, A., Choquet, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeRubeis, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S., Ginde, S., Goyette, M., Graham, W., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., Labocque, K., Lamazares, R., Landers, T., Lebeck, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M., McKean, P., McKernan, K., McSheeters, R., Melchior, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., | | | |

| | | | | |
|-----------------------|--------------|---------------|-----------------|----------------|
| Query Match | 10.6% | Score 51.2; | DB 8; | Length 196859; |
| Best Local Similarity | 45.3%; | Pred. No.1.1; | | |
| Matches 185; | Conservative | 0; | Mismatches 223; | Indels 0; |
| | | | Gaps | 0; |

| | |
|------------|---|
| RESULT 14 | |
| AC124696 | |
| LOCUS | AC124696 |
| DEFINITION | Mus musculus chromosome 16 clone RP24-1731i3, WORKING DRAFT |
| SEQUENCE | 8 unordered pieces. |
| ACCESSION | AC124696 |
| VERSION | AC124696.4 GI:62955023 |
| KEYWORDS | HTG, HTGS_PHASE1, HTGS_DRAFT, HTGS_ACTIVEFIN. |
| SOURCE | Mus musculus (house mouse) |
| ORGANISM | Mus musculus |

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu>
 Contact: submissions@wustl.edu
 ----- Project Information -----
 Center project name: W_BB073113

```

Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator; Big Dye; 100% of reads
Assembly program: PIRAP; version 0.990319
Consensus quality: 191374 bases at least Q40
Consensus quality: 192113 bases at least Q30
Consensus quality: 192485 bases at least Q20

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

| | | | |
|---|-------|---------|-------------------------------|
| * | 1 | 1052: | contig of 1052 bp in length |
| * | 1053 | 1152: | gap of unknown length |
| * | 1153 | 2247: | contig of 1095 bp in length |
| * | 2248 | 2347: | gap of unknown length |
| * | 2348 | 3388: | contig of 1041 bp in length |
| * | 3389 | 3488: | gap of unknown length |
| * | 3489 | 4496: | contig of 1508 bp in length |
| * | 4997 | 5096: | gap of unknown length |
| * | 5097 | 6623: | contig of 1557 bp in length |
| * | 6624 | 6723: | gap of unknown length |
| * | 6724 | 9454: | contig of 2711 bp in length |
| * | 9455 | 9554: | gap of unknown length |
| * | 9555 | 80258: | contig of 70704 bp in length |
| * | 80259 | 193848: | gap of unknown length |
| * | 80359 | 193849: | contig of 113491 bp in length |

```

#RES
source      Location/Qualifiers
1. 193849   /organism="Mus musculus"
           /mol_type="genomic DNA"
           /db_xref="taxon:10090"
           /chromosome="16"
           /clone="RP24-173113"
misc_feature 1. 1052
           /note="assembly_name:Contig23"
gap          1053..1152
           /estimated_length=unknown
misc_feature 1153..2247
           /note="assembly_name:Contig30"
gap          2248..2347
           /estimated_length=unknown

```

| | | | |
|--------------------------|---|---|--------------------------------|
| | misc_feature | 2348..3388 | /note="assembly_name:Contig42" |
| | gap | 3389..3488 | /estimated_length=unknown |
| | misc_feature | 3489..4996 | /note="assembly_name:Contig43" |
| | gap | 4997..5096 | /estimated_length=unknown |
| | misc_feature | 5097..6623 | /note="assembly_name:Contig45" |
| | gap | 6624..6723 | /estimated_length=unknown |
| | misc_feature | 6724..9454 | /note="assembly_name:Contig46" |
| | gap | 9455..9554 | /estimated_length=unknown |
| | misc_feature | 9555..80258 | /note="assembly_name:Contig47" |
| | gap | 80259..80358 | /estimated_length=unknown |
| | misc_feature | 80359..193849 | /note="assembly_name:Contig48" |
| ORIGIN | | | |
| Query Match | | 10.4%; Score 50.4; DB 14; Length 193849; | |
| Best Local Similarity | 56.7%; | Pred. No. 1.6; | |
| Matches 93; Conservative | 0; Mismatches 71; Indels 0; Gaps 0; | | |
| Oy | 43 | TATGAAATATTTTACTGTGATCTTAACAAGTGATATATGGTTGTGTGAATTATTT | 102 |
| Dd | 84521 | TATATATATATATATATATACATCAATCAATCATATTAATGATATATATATGAACATATAT | 84580 |
| Oy | 103 | ATTTGAGAGGATTTTAAACACACTTTCGAATAAATTAATTAATTAATTTCTCTAT | 162 |
| Dd | 84581 | ATTATAGTAATATATATACATATATTTTGAACAACAATTTAATACATACATACATATGAT | 84640 |
| Oy | 163 | CTTTAAGGCACATATTACCGTGCTAAGGCAATACAGCTGAT | 206 |
| Dd | 84641 | ATGATATGTACACATGCCAGATGGAGTTTACATTTGAGCTGACA | 84684 |
| RESULT 15 | | | |
| LOCUS | AC163463 | 244638 bp | DNA linear HTG 01-JUL-2005 |
| DEFINITION | Bos taurus clone CH240-12SK20, *** SEQUENCING IN PROGRESS ***, 39 | | |
| ACCSSION | AC163463 | | |
| VERSION | AC163463.2 GI:68300099 | | |
| KEYWORDS | HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED. | | |
| SOURCE | Bos taurus (cow) | | |
| ORGANISM | Bos taurus | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovine; Bos. | | |
| AUTHORS | 1 (bases 1 to 244638) Muzny,D.,Marie., Metzker,M.,Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Albrooks,S., Amin,A., Angiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalto,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhatz,B., Burck,P., Butrell,K., Calderon,B., Cardenas,V., Carter,K., Cavazos,I., Cesari,H., Centen,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denison,S., Detamo,C., Ding,Y., Dinh,H., Diya,K., Dreier,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregorgis,B., Geer,K., Gill,R., Grady,M., Guerra,W., Guayana,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,J., Harvey,Y., Haylak,P., Hawes,A., Henderson,N., Hernandez,J., | | |

| TITLE | JOURNAL | REFERENCE | AUTHORS | JOURNAL |
|---|---------|-----------|---------|---------|
| Unpublished | | | | |
| 2 (bases 1 to 244638) | | | | |
| Worley,K.C. | | | | |
| Direct Submission | | | | |
| Submitted (12-JUN-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | |
| 3 (bases 1 to 244638) | | | | |
| Cow Genome Sequencing Consortium. | | | | |
| Direct Submission | | | | |
| Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | |
| On Jun 29, 2005 this sequence version replaced gi:67514619. | | | | |
| The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table. | | | | |
| ----- Genome Center | | | | |
| Center: Baylor College of Medicine | | | | |
| Center code: BCM | | | | |
| Web site: http://www.hgsc.bcm.tmc.edu/ | | | | |
| Contact: hgsc-help@bcm.tmc.edu | | | | |
| ----- Project Information | | | | |
| Center project name: FHXA | | | | |
| Center clone name: CH240-125K20 | | | | |
| ----- Summary Statistics | | | | |
| Assembly program: Atlas 3.0/ | | | | |
| Consensus quality: 229556 bases at least Q40 | | | | |
| Consensus quality: 233265 bases at least Q30 | | | | |
| Consensus quality: 236194 bases at least Q20 | | | | |
| Bedimmed insert size: 239301; sum-of-contigs estimation | | | | |
| Quality coverage: 4x in Q20 bases; sum-of-contigs estimation | | | | |

```

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 39 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 4622: contig of 4622 bp in length
* 4623 4672: gap of 50 bp
* 4673 8644: contig of 3972 bp in length
* 8645 8744: gap of unknown length
* 8745 12688: contig of 3944 bp in length
* 12689 12738: gap of 50 bp
* 12739 16581: contig of 3843 bp in length
* 16582 16631: gap of 50 bp
* 16632 22198: contig of 5567 bp in length
* 22199 22248: gap of 50 bp
* 22249 26048: contig of 3800 bp in length
* 26049 26098: gap of 50 bp
* 26099 37514: contig of 11416 bp in length
* 37515 38190: gap of 676 bp
* 38191 55334: contig of 17144 bp in length
* 55335 55384: gap of 50 bp
* 55385 75854: contig of 20470 bp in length
* 75855 76343: gap of 489 bp
* 76344 76344: contig of 1 bp in length
* 76345 76394: gap of 50 bp
* 76395 83875: contig of 7481 bp in length
* 83876 83925: gap of 50 bp
* 83926 94134: contig of 10209 bp in length
* 94135 94267: gap of 133 bp
* 94268 97884: contig of 3517 bp in length
* 97885 97885: gap of 50 bp
* 97886 108699: contig of 10865 bp in length
* 108700 109258: gap of 559 bp
* 109259 115882: contig of 6624 bp in length
* 115883 115932: gap of 50 bp
* 115933 121475: contig of 5543 bp in length
* 121476 121525: gap of 50 bp
* 121526 131673: contig of 10148 bp in length
* 131674 131723: gap of 50 bp
* 131724 140028: contig of 8305 bp in length
* 140029 140128: gap of unknown length
* 140129 151032: contig of 10904 bp in length
* 151033 151082: gap of 50 bp
* 151083 181837: contig of 30755 bp in length
* 181838 181887: gap of 50 bp
* 181888 184514: contig of 2627 bp in length
* 184515 184564: gap of 50 bp
* 184565 189352: contig of 4788 bp in length
* 189353 189402: gap of 50 bp
* 189403 195600: contig of 6198 bp in length
* 195601 195650: gap of 50 bp
* 195651 201941: contig of 6291 bp in length
* 201942 201991: gap of 50 bp
* 201992 207379: contig of 5388 bp in length
* 207380 207429: gap of 50 bp
* 207430 209271: contig of 1842 bp in length
* 209272 209371: gap of unknown length
* 209372 211084: contig of 1713 bp in length
* 211085 211134: gap of 50 bp
* 211135 220592: contig of 9458 bp in length
* 220593 220642: gap of 50 bp
* 220643 223227: contig of 2585 bp in length
* 223228 223327: gap of unknown length
* 223328 224454: contig of 1127 bp in length
* 224455 226141: contig of 1587 bp in length
* 226142 226241: gap of unknown length

```

```

* 226242 227620: contig of 1379 bp in length
* 227621 227720: gap of unknown length
* 227721 229657: contig of 1937 bp in length
* 229658 229757: gap of unknown length
* 229758 231422: contig of 1665 bp in length
* 231423 231522: gap of unknown length
* 231523 233019: contig of 1497 bp in length
* 233020 233119: gap of unknown length
* 233120 235005: contig of 1886 bp in length
* 235006 235105: gap of unknown length
* 235106 236675: contig of 1570 bp in length
* 236676 236775: gap of unknown length
* 236776 240371: contig of 3596 bp in length
* 240372 240471: gap of unknown length
* 240472 244638: contig of 4167 bp in length.
*
FEATURES
    source
        1..244638
            /organism="Bos taurus"
            /mol_type="genomic DNA"
            /db_xref="taxon:9913"
            /clone="CH240-125K20"
            4623..4672
            240471..244638: estimated length=50
            8645..8744
            gap
                10.3%; Score 50; DB 14; Length 244638;
                Best Local Similarity 36.9%; Pred. No. 1.8;
                Matches 152; Conservative 0; Mismatches 260; Indels 0; Gaps 0;

```

Search completed: December 28, 2005, 10:11:47
Job time : 2763 secs

```

QY 39 AAGATGATGAAATTTTCTGCTGATCTTACAGTGTATATGCTGTGCTGACTT 98
DB 76496 AATTTTAAATTTTATTTATTTAAATTTAAATTTTAAATTTTAAATTTT 76555
QY 99 ATTATTTGAGAGTATTTTAAACACCTTAGAATTTAAATTTAAATTTTCT 158
DB 76556 ATTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 76615
QY 159 CTATCTTTAAGGACATTTTACGCTGCTAGGCAATTTACGCTGCTAACT 218
DB 76616 AATATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 76675
QY 219 CATGCGCACATTTCTTCTTAAACGCGTCTGCTTCCAAAGGACCTCGAAT 278
DB 76676 ATACATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 76735
QY 279 GCACTATTTATCTGCGCATTTCCAAATTTATCCCTATTTGGATTTGATG 338
DB 76736 TTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 76795
QY 339 GTTTAATGTCACCGATTTGAATCTTCACTTGTGAGTTTGTCTTGTCTTAA 398
DB 76796 TTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 76855
QY 399 GGTCTTCAATTTATCTTAAAGCAAGTTTGTATTAATTTCAAAATCTTGT 450
DB 76856 TTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 76907

```

this Page Blank (uspto)

XX SQ Sequence 485 BP; 137 A; 90 C; 82 G; 176 T; 0 U; 0 Other;
Query Match 100.0%; Score 485; DB 12; Length 485;
Best Local Similarity 100.0%; Pred. No. 3.3e-104;
Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGCGGAGAGAGAGAAAGTGAAGTCTGTTGCGTGAAGACATGATGAAAAATTTTACT 60
DB 1 GCGCGGAGAGAGAGAAAGTGAAGTCTGTTGCGTGAAGACATGATGAAAAATTTTACT 60
QY 61 GGTGATCTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
DB 61 GGTGATCTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
QY 121 CACACCTTGAAGCTAAATCTAATAATATTTCTATCTTTAAGGACATATTA 180
DB 121 CACACCTTGAAGCTAAATCTAATAATATTTCTATCTTTAAGGACATATTA 180
QY 181 CGTGGCTAAGCAATTACAGCTGATATCTGTAATATCTGATGCGCACTAAATCTTCT 240
DB 181 CGTGGCTAAGCAATTACAGCTGATATCTGTAATATCTGATGCGCACTAAATCTTCT 240
QY 241 AACAGCGCTTCTGCTCTTCCAGGAGCTCCGAATATGCGCACTATTTATCTGCGCAT 300
DB 241 AACAGCGCTTCTGCTCTTCCAGGAGCTCCGAATATGCGCACTATTTATCTGCGCAT 300
QY 301 TCCAAATTTAATCCCTATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
DB 301 TCCAAATTTAATCCCTATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
QY 361 TCTTCACTGTTGAGTTTGTCTTTTGTCTCTTAAGGCTTCAATTTATCTAAGCA 420
DB 361 TCTTCACTGTTGAGTTTGTCTTTTGTCTCTTAAGGCTTCAATTTATCTAAGCA 420
QY 421 AGTTTGTATTAATGAAATATCTTCTTCTTCCATGATGATGATGATGATGATGATGATG 480
DB 421 AGTTTGTATTAATGAAATATCTTCTTCTTCCATGATGATGATGATGATGATGATGATG 480
QY 481 GTACC 485
DB 481 GTACC 485

RESULT 2
AD039660
ID AD039660 standard; DNA; 4046 BP.
XX
AC AD039660;
XX
DT 29-JUL-2004 (first entry)
XX
DE Yeast chitin synthase 3 CAL1 gene.
XX
KW Heterologous gene; expression cassette; gene expression; Yeast; CAL1;
KW chitin synthase 3; gene; ds.
XX
OS Saccharomyces cerevisiae.
XX
PN US2004092020-A1.
XX
PD 13-MAY-2004.
XX
PF 20-JUN-2003; 2003US-00600230.
XX
PR 20-JUN-2002; 2002US-0390529P.
XX
PA (WILK/) WILKINSON J Q.
XX
PA (MCBR/) MCBRIDE K.
XX
PA (BERT/) BERTAIN S.
XX
PI Wilkinson JQ, McBride K, Bertain S;
XX

DR WPI; 2004-374960/35.
DR GENBANK; X57300.
XX
XX PT New recombinant expression cassette comprising a promoter that is
XX functional in plants, operably linked with a coding sequence and a non-
XX plant 3' termination sequence, useful for gene expression in plant cells.
XX
XX PS Disclosure; SEQ ID NO 81; 74bp; English.
XX
XX CC The present invention relates to heterologous genes comprising non-plant
XX 3'-terminal sequences and plant expression cassettes incorporating the
XX heterologous genes. The invention is useful for gene expression in plant
XX cells. The present sequence is yeast chitin synthase 3 CAL1 gene. This
XX sequence is used in the invention.
XX

QY Sequence 4046 BP; 1211 A; 739 C; 819 G; 1277 T; 0 U; 0 Other;
Query Match 97.9%; Score 474.8; DB 12; Length 4046;
Best Local Similarity 99.6%; Pred. No. 1.2e-101;
Matches 476; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 CGAAGAGAGAAAGTGAAGTCTGTTGCGTGAAGACATGATGAAAAATTTTACTGAT 65
DB 3546 CGAAGAGAGAAAGTGAAGTCTGTTGCGTGAAGACATGATGAAAAATTTTACTGAT 3605
QY 66 ACTTCAAGTATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 125
DB 3606 ACTTCAAGTATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3665
QY 126 CTTAAGCTAAATCTAATAATTAATTTCTATCTTTAAGGACATATTAAGTGG 185
DB 3666 CTTAAGCTAAATCTAATAATTAATTTCTATCTTTAAGGACATATTAAGTGG 3725
QY 186 CTAAAGCAATTAAGCTATATCTGTAATATCTGTAATATCTGTAATATCTGTAATATCTG 245
DB 3726 CTAAAGCAATTAAGCTATATCTGTAATATCTGTAATATCTGTAATATCTGTAATATCTG 3785
QY 246 GCGTCTGCTCTTCCAGGAGCTCCGAATATGCGCACTATTTATCTGCGCATTTCCAA 305
DB 3786 GCGTCTGCTCTTCCAGGAGCTCCGAATATGCGCACTATTTATCTGCGCATTTCCAA 3845
QY 306 TTTATATCCCTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 365
DB 3846 TTTATATCCCTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3905
QY 366 ACTTGTGAGTTTGTCTTTTGTCTCTTAAGGCTTCAATTTATCTAAGCAAGTTT 425
DB 3906 ACTTGTGAGTTTGTCTTTTGTCTCTTAAGGCTTCAATTTATCTAAGCAAGTTT 3965
QY 426 TGTATATTCAAATATCTTGTCTTCTTCCATGATGATGATGATGATGATGATGATGATG 483
DB 3966 TGTATATTCAAATATCTTGTCTTCTTCCATGATGATGATGATGATGATGATGATGATG 4023

RESULT 3
ABL32718
ID ABL32718 standard; DNA; 5739 BP.
XX
AC ABL32718;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 691.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antihaemic; cytoseric; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antidiabetic; antidiabetic; antipsoriasis;
KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW de.
XX

OS Homo sapiens.
 XX
 XX WO200200928-A2.
 XX
 XX
 PD 03-JAN-2002.
 PF 02-JUL-2001; 2001WO-EP007537.
 XX
 XX 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 XX (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR MPI; 2002-130909/17.
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 XX
 PS Claim 1; SEQ ID NO 691; 32pp + Sequence Listing; German.
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 CC
 SQ Sequence 5739 BP; 1457 A; 38 C; 1217 G; 3027 T; 0 U; 0 Other;
 Query Match 9.5%; Score 46.2; DB 6; Length 5739;
 Best Local Similarity 45.9%; Pred. No. 0.5; Mismatches 228; Indels 1; Gaps 1;
 Matches 194; Conservative 0;
 Db 31 TTGCGTAGACAGATGAGAAATATTTTACTGATGATCTTCAAGTATATGATG 90
 729 TTGATTAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 788
 Qy 91 TGTAACTATTTTATTTGAGAGTATTTTAAACACCTTGAAGTAAATTAATTA 150
 789 TTTTAGTTTATTTTGAATTTTGAATTTTGAAGAAAGTTTAAAT-AGTTTTCATGTT 847
 Db 151 ATATTTCTATCTTTTAAAGCACATTTACGTGGCTTAAGCAATTCACCTGATAT 210
 848 ATATTTTAAATTAATTTTGTGAATTTTGTGAATTTTAAATTAATTAATTTGTTGA 907
 Qy 211 GTAAACCTCAGTGGCAGTAAATTTCTCTAAACAGCGTCTGCTCTTCAAGGACT 270
 908 TTAATTTTATTTTGAAGATTTGTTTATTAATTTTGTGTTTATTTTATTTTATTTT 967
 Db 271 CCGAATATGCACTATTTATCTGTGCAATTTCCAAATTAATTCCTATTTGGTAT 330
 968 TTTTATTAATTAATTTTGTGTTATTTTGTGTTATTTTGTGTTTAAATTAATTTGATG 1027
 Qy 331 ATGTGGCCGTTAAATAGTCAAGCATTTCTTCACTGTTGCAAGTTTGTCTTTGCT 390
 1028 TTAAGTTTGTATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1087
 Db 391 TCTCTAAAGTCTCAATTTATCTAAAGCAAGTTTGTATTAATCAAAATATCTTTGCTT 450
 1088 TTTTATATGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1147
 Qy 451 TCT 453
 Db 1148 TAT 1150

ID AAS61080 standard; DNA; 6092 BP.
 XX
 XX AAS61080;
 XX
 XX
 DT 29-JAN-2002 (first entry)
 XX
 XX
 DE Human gene regulation-associated gene oligonucleotide #35.
 XX
 XX Human; Gene regulation-associated gene; severe combined immunodeficiency;
 XX cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
 XX asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
 XX renal disease; Pre-eclampsia; cardiac allograft vascular disease;
 XX colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
 XX immunostimulant; cardiac; anti-inflammatory; coagulant; antistatic;
 XX nephrotropic; gynecological; anti-tumour; immunosuppressive; cyostatic.
 OS Homo sapiens.
 XX
 XX WO200177375-A2.
 XX
 PD 18-OCT-2001.
 PF 06-APR-2001; 2001WO-EP003968.
 XX
 XX
 PR 06-APR-2000; 2000DE-01019058.
 PR 07-APR-2000; 2000DE-01019173.
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 XX (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR MPI; 2002-017470/02.
 XX
 PT New nucleic acid sequences from chemically modified genes associated with
 PT gene regulation, useful for analyzing cytosine methylations for diagnosis
 PT and therapy of diseases e.g. severe combined immunodeficiency disease.
 XX
 PS Claim 1; SEQ ID NO 36; 26pp; English.
 CC The invention relates to 224 nucleic acid sequences comprising at least
 CC 18 bases of a chemically pretreated gene associated with gene regulation
 CC selected from 43 known genes (or complementary sequences). The chemical
 CC pretreatment converts cytosine bases unmethylated at the 5-position to
 CC uracil or another base with hybridisation behaviour dissimilar to
 CC cytosine, to enable analysis of cytosine methylations. The DNA sequences,
 CC oligomers (or sets/arrays) and method are useful in the diagnosis of
 CC diseases (or predisposition to diseases) associated with gene regulation
 CC and in therapy of such diseases, by enabling analysis of the cytosine
 CC methylation patterns of such genes. Kits are provided. They are
 CC especially useful in diagnosis and therapy of e.g. severe combined
 CC immunodeficiency disease, cardiac disorders, haemophilia, solid tumours
 CC and cancer, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen
 CC syndrome, renal disease, pre-eclampsia, graft versus-host disease. The
 CC present sequence is a sequence included in the sequence data for this
 CC specification and is associated with the human gene regulation-associated
 CC genes. Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from Wipo at http://wipo.int/pub/published_pct_sequences
 CC
 SQ Sequence 6092 BP; 1931 A; 122 C; 1140 G; 2899 T; 0 U; 0 Other;
 Query Match 9.5%; Score 46.2; DB 6; Length 6092;
 Best Local Similarity 46.8%; Pred. No. 0.5; Mismatches 203; Indels 3; Gaps 1;
 Matches 181; Conservative 0;
 Qy 62 TGAATCTTACCAAGTGAATATATGTTGTGTAATTTATTTATTTTAAAGTATTTTAAAC 121
 4944 TGTATTTTAAAGTGTGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 5003
 Db 122 ACACCTTAGACATTAATTAATTAATTAATTTCTCATCTTTAAAGCACATATTC 181

Db 5004 ATTTGTTTAAATATTAATTGCGAAATATTAGTTTTTATTATTAAAGTTTGAAATTAT 5063
 QY 182 GTGGCTAAGGCAATTACAGCTGATATACGTAAAACTCATGTGCGACATAATCTCTCA 241
 Db 5064 AATATGATATTAATATAATATGATATATTTTATTAAATTTGTTTTAAATTTTAA 5123
 QY 242 ACAAGCGTTCTGCTCTCTTCCAAAGGACCTCCGAATATGCCACTATTTATCTGGCAATT 301
 Db 5124 TGTGTGATTTTATATATTTATATAGTAAATAAATATGA---TAAATTAATTTGTGGAAT 5180
 QY 302 CCAATTTATATTCCTCATTTGGGATTTGATGAGCGCTTAAATAGTACCGATTTGAAT 361
 Db 5181 ATAGTTTATGTTTAAAGTTTATAGATTAATTTATTTATATATGTTTGTATATA 5240
 QY 362 CTTCATCTTGTCGAGTTTGTCTTTGCTCTCTTAAGGCTCTCAATTTATCTTAAAGCA 421
 Db 5241 TATAGTATTAATTTATATAGTAAATTTATTTTAAATATTTGATTTGATTTAAAGTAA 5300
 QY 422 GTTTGTATTAATTCAAAATATCTTGCT 448
 Db 5301 AGAATTAATATTAATAAGTATTTGAT 5327

RESULT 5

AB066998 standard; DNA; 37515 BP.

AB066998;

28-AUG-2002 (first entry)

Human angiogenesis associated polynucleotide SEQ ID NO 28.

Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
 inflammation; rheumatoid arthritis; diabetic retinopathy; antileucers;
 macular degeneration; inflammatory bowel disease; Crohn's disease;
 anti-neumatic; antidiabetic; antidiabetic; antipsoriatic;
 antileucers; ds.

Homo sapiens.

WO20024654-A2.

13-JUN-2002.

06-DEC-2001; 2001WO-BP014320.

06-DEC-2000; 2000DE-01061338.

(EPIC-) EPIGENOMICS AG.

Schacht O;

WPI; 2002-500450/53.

New nucleic acid fragments from chemically treated angiogenesis-
 associated genes, useful for determining methylation status, e.g. in
 diagnosis or treatment of cancer.

Claim 1; SEQ ID NO 28; 41bp + Sequence Listing; German.

The invention relates to a nucleic acid (I) comprising a segment of 18
 bases of chemically pretreated DNA of angiogenesis-associated genes (II)
 having sequences (AB066997-AB067178) or their complements. (I), also
 related oligomers, are used to evaluate the methylation status and/or
 single-nucleotide polymorphisms, in angiogenesis-related genes, for
 diagnosis and treatment of eye diseases, proliferative retinopathy,
 neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
 diabetic retinopathy, macular degeneration caused by neovascularisation,
 psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
 Crohn's disease. Note: The sequence data for this patent did not form
 part of the printed specification, but was obtained in electronic format
 directly from WIPO at ftp.wipo.int/pub/published_pcc_sequences

SX Sequence 37515 BP; 11428 A; 425 C; 7165 G; 18497 T; 0 U; 0 Other;
 SQ Query Match 9.3%; Score 45; DB 6; Length 37515;
 Best Local Similarity 53.0%; Pred. No. 1.3;
 Matches 96; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 43 TATGAAATATTTTACTGATACCTTACAGGTTGATATGTTGTGTTAACTTATTT 102
 Db 14019 TATATATTTTATATTTATTAATTAATTTTATTAATTTTGTATGATGAATTT 14078
 QY 103 ATTTGAGGATTTTAAACACCTTGAACCTTAACTTAATTAATTAATTTCTAT 162
 Db 14079 GTTTGTGATTAATTAATTAATTTTATTAATTTTATTAATTTATTTGTTTGT 14138
 QY 163 CTTTAAAGCAGATTAATGCTGCTAAGGCAATTACAGCTGATATCTGTAATAATG 222
 Db 14139 AATTTTGTATTTTTCGTAAGTTTAAAGAAATGAATGAGTATGATATTAATG 14198
 QY 223 T 223
 Db 14199 T 14199

RESULT 6

ABL32336 standard; DNA; 7201 BP.

ABL32336;

26-MAR-2002 (first entry)

Human immune system associated gene SEQ ID NO: 309.

Human; immune system disease; cytosine methylation; antileukemic;
 antileukemic; antileukemic; cytosine methylation; antileukemic;
 antileukemic; anti-HIV; anticonvulsant; ophthalmological;
 neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 anti-inflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 ds.

Homo sapiens.

WO200200928-A2.

03-JAN-2002.

02-JUL-2001; 2001WO-BP007537.

30-JUN-2000; 2000DE-01032529.

01-SEP-2000; 2000DE-01043826.

(EPIC-) EPIGENOMICS AG.

Olek A, Piepenbrock C, Berlin K;

WPI; 2002-130909/17.

Nucleic acid comprising fragment of chemically modified gene, useful for
 diagnosis and treatment of diseases associated with abnormal cytosine
 methylation.

Claim 1; SEQ ID NO 309; 32bp + Sequence Listing; German.

The present invention provides a number of human immune system associated
 genes which are modified by the methylation of cytosines. The sequences
 can be used in the diagnosis and treatment of immune system disorders,
 including eye diseases such as retinopathy, neovascular glaucoma and
 macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel


```
RESULT 10
ACL37108/c
ID ACL37108 standard; cDNA; 2000 BP.
XX
AC ACL37108;
XX
DT 02-JUN-2005 (first entry)
XX
DE Rice stress-regulated promoter SEQ ID NO:15671.
XX
KM ss; abiotic stress tolerance; transgenic plant; plant; cereal;
XX agriculture.
XX
OS Oryza sativa.
XX
PN MO2003008540-A2.
XX
PD 30-JUN-2003.
XX
PF 21-JUN-2002; 2002MO-US019668.
XX
PR 22-JUN-2001; 2001US-0300112P.
XX
PR 24-AUG-2001; 2001US-0314662P.
XX
PR 26-SEP-2001; 2001US-0325277P.
XX
PR 21-NOV-2001; 2001US-0332132P.
XX
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI Krepe J, Briggs SP, Cooper B, Glazebrook J, Goff SA, Katagiri F,
PI Moughamer T, Frouart N, Rieke D, Zhu T;
XX
DR WPI; 2003-248011/24.
XX
PT New stress-responsive nucleic acid, useful for altering the
PT responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
PT stresses, salt stress or osmotic stress.
XX
PS Claim 48; SEQ ID NO 15671; 89bp; English.
XX
XX
CC The invention relates to novel abiotic stress responsive polynucleotides
CC and polypeptides. Also disclosed are vectors, expression cassettes, host
CC cells, and plants containing such polynucleotides. Also disclosed are
CC methods for using the polynucleotides and polypeptides to alter the
CC responsiveness of a plant to abiotic stress. The invention is useful in
CC agriculture. The nucleic acid is useful for determining whether a test
CC plant has been exposed to an abiotic stress condition. It is also useful
CC for selecting an agent that alters abiotic stress regulated
CC polynucleotide expression in a plant cell, and to identify a homolog or
CC ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
CC molecule and the polypeptide encoded by it are useful in altering the
CC responsiveness of a plant to an abiotic stress, such as cold stress, salt
CC stress, osmotic stress or any of their combinations. The present sequence
CC is used in the exemplification of the invention
XX
SQ Sequence 2000 BP; 510 A; 350 C; 268 G; 518 T; 0 U; 354 Other;
XX
Query Match 8.9%; Score 43; DB 11; Length 2000;
Best local Similarity 12.2%; Pred. No. 2.3; Mismatches 4; Gaps 2;
Matches 55; Conservative 206; Mismatches 186; Indels 4; Gaps 2;
XX
QY 9 AAGAGGAAAGTACTCTTCGTTGCGTAGACAGTATGAAATATTTTACTGTGATACT 68
DB 451 ARKAWGKARWVCKRTTAAVATMTTTRRRRAAMWMTGKRKRBSMTTRMTSRMTTWGM 392
QY 69 TACAATTTGATATNTGTTGTGTGTAACCTTA--TTTATTTGAGAGGATATTTAACACACC 126
DB 391 YKYYTWTTRRRYSRRRTGKAKKYYMTWMMWGCMSYTKARWMAAYKSWMAVYAWT 332
QY 127 TTGACACTAAACTATATATATATATATTTCTATCTTAAAGGACATATTAAGCTGGC 186
DB 331 KGTRRYYMKRAAYVTRMTATTSATWCTGWTYTSKATTTKMGAGVAKATTTMKRKRY 272
```

```
QY 187 TAAGCAATTACAGCTGATATATACGTAAACTCATGTCGCCACTAAATTTCTTACACAG 246
DB 271 YAGGRWWWYKKWASRTWKMCMYKMYMNAATKSSMWSMATYAMKSMWYKKTYYTMTS 212
QY 247 CGTTCGTCTCTTCCCAAGGAGCTCCGAATATNGCCAC--ATTATTCGTGGCATTTCCA 304
DB 211 TYWTWMMWRCPAGWMMWGGAKRAAATKMYRTYRKMMRYRYKMSWAGMAMRKR 152
QY 305 ATTATATATTCCTTATTTGGGATTTGATGAGCCGCTTAAATAGTACGATGATCTT 364
DB 151 YTTWMMWMMYRCAMWMSAYWRSWRMTTTKSKMYKMMWMTTYABAKAYSRMMWK 92
QY 365 CACTGTTCGAGTTTGTCTTTGCTTCTCTTAAAGGTCTTCAATTATTAAGCAAGTT 424
DB 91 KRCRTRMKGWTRKTKWMMWRWYKRYRKMSKATWYRSGAYRRBWRBMAYSRRSW 32
QY 425 TTGTATTAATTCAAATTAATCTTCTTCTCC 455
DB 31 MKRTSYYYTKWKKWKMSSGYWRYYTTMS 1
XX
RESULT 11
ADZ71063/c
ID ADZ71063 standard; DNA; 847 BP.
XX
AC ADZ71063;
XX
DT 14-JUL-2005 (first entry)
XX
DE Human chromosome 2 contig DNA SEQ ID NO 187.
XX
KM ds; matrix attachment region; MAR; protein production.
XX
OS Homo sapiens.
XX
PN WO2005040377-A2.
XX
PD 06-MAY-2005.
XX
PF 22-OCT-2004; 2004MO-EP011974.
XX
PR 24-OCT-2003; 2003US-0513574P.
XX
PR 06-FEB-2004; 2004EP-00002722.
XX
PA (SELEX-) SELEXIS SA.
XX
PI Mermoud N, Girard PA, Bucher P, Nguyen D, Calabrese D, Sauvy D;
PI Puttini S;
XX
DR WPI; 2005-333507/34.
XX
PT New purified and isolated DNA sequence having protein production
PT increasing activity comprises a bent DNA element and a binding site for a
PT DNA binding protein, useful for increasing protein production activity in
PT eukaryotic host cell.
XX
PS Disclosure; SEQ ID NO 187; 282bp; English.
XX
XX
CC The invention relates to a purified and isolated DNA sequence having
CC protein production increasing activity comprising at least one bent DNA
CC element, and at least one binding site for a DNA binding protein. The
CC purified and isolated DNA sequence comprising a first and a second
CC isolated matrix attachment region (MAR) nucleotide sequence, which is a
CC MAR nucleotide sequence selected from a purified and isolated DNA
CC sequence above, a purified and isolated MAR DNA above, a purified and
CC isolated clysmar element and/or fragment, a synthetic MAR sequence, a
CC sequence complementary to it, its molecular chimera, or its combinations
CC and variants, is useful for increasing protein production activity in a
CC eukaryotic host cell. The present sequence represents a human chromosome
CC 2 contig DNA.
XX
SQ Sequence 847 BP; 501 A; 17 C; 10 G; 319 T; 0 U; 0 Other;
```



```

FT      /tag= af
FT      /number= 11
FT      exon      19566..19628
FT      /tag= ag
FT      /number= 12
FT      intron    19629..19885
FT      /tag= ah
FT      /number= 12
FT      exon      19886..19950
FT      /tag= ai
FT      /number= 13
FT      intron    19951..20580
FT      /tag= aj
FT      /number= 13
FT      exon      20581..20748
FT      /tag= ak
FT      /number= 14
FT      variation replace(20901,T)
FT      /tag= al
FT      /standard_name= "Single nucleotide polymorphism"
FT      variation replace(21098,G)
FT      /tag= am
FT      /standard_name= "Single nucleotide polymorphism"
FT      variation replace(21101,G)
FT      /tag= an
FT      /standard_name= "Single nucleotide polymorphism"
PN      WO200194363-A2.
XX      13-DEC-2001.
XX      07-JUN-2001; 2001WO-US018458.
XX      07-JUN-2000; 2000US-0209935P.
XX      (GENA-) GENAISSANCE PHARM INC.
XX      Duda A, Kazemi A, Koshy B;
XX      WPI: 2002-566434/60.
XX      P-PSDB; ABP57966.
XX      New 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 2 (PFKFB2) gene
XX      variants, for improving efficiency and reliability in the development of
XX      drugs for treating diseases associated with PFKFB2 activity e.g. cancer.
XX      Example 1; Fig 1; 95pp; English.
XX      The invention relates to a novel human 6-phosphofructo-2-kinase/ fructose
XX      -2,6-bisphosphatase 2 (PFKFB2) isogene. The PFKFB2 of the invention has
XX      cytosolic and antidiabetic activity. The polymucleotides may have a use
XX      in gene therapy. The identified candidate agents targeting PFKFB2, are
XX      useful for treating cancer and diabetes. The methods of the invention are
XX      useful for improving the efficiency and reliability of several steps in
XX      the discovery and development of drugs for treating diseases associated
XX      with PFKFB2 activity. The present sequence encodes the PFKFB2 of the
XX      invention
XX      Sequence 22617 BP; 5587 A; 4922 C; 5366 G; 6741 T; 0 U; 1 Other;
SQ      Query Match      8.7%; Score 42.2; DB 6; Length 22617;
      Best Local Similarity 57.0%; Pred. No. 5.5;
      Matches 77; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY      43 TATGAATAATTTTACTGATGATCTTACAAAGTTGATATGCTGTGCTAATTT 102
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      9535 TATATAAGATATATATAATATATATATATATATATATATATATATATAT 9476
QY      103 ATTGAGAGGATATTTTAAACACCTTAGAAGTAACTAAATTAATATTTCTCTAT 162
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      9475 ATATTTAAAGATATTTATATATATCTTAAATATATATATTTAAGATATATATATAG 9416
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      163 CTTTAAAGGACATTA 177

```

```

DB      9415 ATATATAGATATATA 9401
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 13
ID      AAS46704
ID      AAS46704 standard; DNA; 14615 BP.
XX      AAS46704;
XX      18-DEC-2001 (first entry)
XX      Tumour suppressor gene derived chemically modified sequence #427.
XX      DE
XX      Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
XX      KM tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
XX      KM cytosine methylation; de.
XX      OS Homo sapiens.
XX      OS
XX      WO200168912-A2.
XX      PD 20-SRP-2001.
XX      PF 15-MAR-2001; 2001WO-EP002955.
XX      PR 15-MAR-2000; 2000DE-01013847.
XX      PR 06-APR-2000; 2000DE-01019058.
XX      PR 07-APR-2000; 2000DE-01019173.
XX      PR 30-JUN-2000; 2000DE-01032529.
XX      PR 01-SEP-2000; 2000DE-01043826.
XX      PA (EPIG-) EPIGENOMICS AG.
XX      PI Olek A, Piepenbrock C, Berlin K;
XX      DR WPI; 2001-602752/68.
XX      PT Fragments of chemically modified genes associated with tumor suppressor
XX      PT genes and oncogenes, useful in designing primers and probes for analyzing
XX      PT diseases associated with cytosine methylation state e.g. cancer.
XX      PS Claim 1; SEQ ID NO 427; 27pp; English.
XX      CC The invention relates to a nucleic acid comprising a sequence of 18
XX      CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
XX      CC bisulphite, of genes associated with tumour suppression and oncogenes
XX      CC and having a sequence taken from 536 (actually 533 since numbers 408, 458 and
XX      CC 500 are missing from the sequence listing) sequences (S8) and sequences
XX      CC complementary to (S8). The nucleic acid may be a peptide nucleic acid-
XX      CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
XX      CC probes for detecting the cytosine methylation state and/or single
XX      CC nucleotide polymorphisms and also to be used in an array for analysing
XX      CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
XX      CC probes can also be used in a method for ascertaining genetic and/or
XX      CC epigenetic parameters for the diagnosis and/or therapy of existing
XX      CC diseases or the predisposition to specific diseases, by analysing
XX      CC cytosine methylations. The parameters may be compared to another set of
XX      CC genetic and/or epigenetic parameters, which are disadvantageous to
XX      CC for diagnosis and/or prognosis events, which are disadvantageous to
XX      CC patients. The present sequence is one of the 533 genomic sequences
XX      CC derived from tumour suppressor genes and oncogenes. Note: The sequence
XX      CC data for this patent did not form part of the printed specification, but
XX      CC was obtained in electronic format directly from WIPO at
XX      CC ftp.wipo.int/pub/published_pct_sequences
SQ      Sequence 14615 BP; 4072 A; 253 C; 3233 G; 7057 T; 0 U; 0 Other;
QY      Query Match      8.7%; Score 42; DB 4; Length 14615;
      Best Local Similarity 56.5%; Pred. No. 5.6;
      Matches 78; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY      43 TATGAATAATTTTACTGATGATCTTACAAAGTTGATATATGCTGTGCTAATTT 102

```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 28, 2005, 08:09:16 ; Search time 3840 Seconds
(without alignments)
5909.306 Million cell updates/sec

Title: US-10-600-230-1

Sequence: 1 gcgcgcgcgaaggaggaagt.....cctccaatgatgaggtacc 485

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database:

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_hlc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gsa1:*
10: gb_gsa2:*
11: gb_gsa3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 66.8 | 13.8 | 495 | 9 | BZ302050 |
| 2 | 55.6 | 11.5 | 1001 | 10 | CNS0064G |
| 3 | 51.8 | 10.7 | 1101 | 10 | CNS0039G |
| 4 | 48.8 | 10.1 | 1101 | 10 | CNS006VL |
| 5 | 47.8 | 9.9 | 1191 | 10 | CL104745 |
| 6 | 47.6 | 9.8 | 705 | 9 | BH945927 |
| 7 | 47 | 9.7 | 1101 | 10 | CNS0106X |
| 8 | 46.8 | 9.6 | 1101 | 10 | CNS016LT |
| 9 | 46.2 | 9.5 | 1101 | 10 | CNS0182P |
| 10 | 46.2 | 9.5 | 1161 | 1 | AJ929051 |
| 11 | 45.8 | 9.4 | 1079 | 1 | AJ928310 |
| 12 | 45.4 | 9.4 | 1014 | 10 | DU000686 |
| 13 | 45.2 | 9.3 | 698 | 8 | CX173493 |
| 14 | 45.2 | 9.3 | 718 | 9 | CE216145 |
| 15 | 45 | 9.3 | 215 | 1 | AV112242 |
| 16 | 45 | 9.3 | 1484 | 9 | CC229324 |
| 17 | 44.6 | 9.2 | 717 | 9 | BZ454798 |
| 18 | 44.6 | 9.2 | 891 | 10 | CZ977617 |
| 19 | 44.6 | 9.2 | 914 | 10 | CNS002JY |
| 20 | 44.6 | 9.2 | 1101 | 10 | CNS001PB |
| 21 | 44.4 | 9.2 | 974 | 10 | CNS001TT |
| 22 | 44 | 9.1 | 1204 | 10 | CNS016E2 |

ALIGNMENTS

| | | | | | |
|----|------|-----|------|----|----------|
| 23 | 43.8 | 9.0 | 928 | 10 | CNS00DXY |
| 24 | 43.6 | 9.0 | 549 | 6 | CD441919 |
| 25 | 43.6 | 9.0 | 733 | 10 | CE692306 |
| 26 | 43.6 | 9.0 | 868 | 10 | CNS010YV |
| 27 | 43.6 | 9.0 | 966 | 6 | CD388381 |
| 28 | 43.4 | 8.9 | 332 | 3 | BI814443 |
| 29 | 43.4 | 8.9 | 921 | 10 | AG552152 |
| 30 | 43.4 | 8.9 | 987 | 10 | CNS014PQ |
| 31 | 43.4 | 8.9 | 1057 | 8 | DN782797 |
| 32 | 43.4 | 8.9 | 1101 | 10 | CNS0039R |
| 33 | 43.2 | 8.9 | 547 | 3 | BM535026 |
| 34 | 43.2 | 8.9 | 555 | 9 | AO451200 |
| 35 | 43.2 | 8.9 | 750 | 10 | CNS011TD |
| 36 | 43.2 | 8.9 | 811 | 1 | AL514901 |
| 37 | 43.2 | 8.9 | 1258 | 10 | CL508886 |
| 38 | 43 | 8.9 | 434 | 9 | BH183920 |
| 39 | 43 | 8.9 | 434 | 11 | CNS070Z8 |
| 40 | 43 | 8.9 | 879 | 10 | CL486692 |
| 41 | 43 | 8.9 | 1126 | 9 | CC215864 |
| 42 | 43 | 8.9 | 1203 | 10 | CNS015WU |
| 43 | 42.8 | 8.8 | 415 | 9 | AQ203034 |
| 44 | 42.8 | 8.8 | 422 | 3 | BI815751 |
| 45 | 42.8 | 8.8 | 697 | 9 | BH700251 |

RESULT 1
LOCUS BZ302050 495 bp DNA linear GSS 31-OCT-2002
DEFINITION K10625.p1 Kluveromyces delphensis Random Genomic Library
Kluveromyces delphensis genomic clone K10625, genomic survey
sequence.

ACCESSION BZ302050.1 GI:24448520
VERSION BZ302050
SOURCE GSS.
Kluveromyces delphensis
Kluveromyces delphensis
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Kluveromyces.

REFERENCE 1 (bases 1 to 495)
Wong,S., Fares,M.A., Zimmermann,W., Butler,G. and Wolfe,K.H.
Evidence from comparative genomics for a complete sexual cycle in
the 'asexual' pathogenic yeast *Candida glabrata*
Genome Biol. 4 (2), R10 (2003)

JOURNAL PUBLISHED 12620120
Contact: Wong S
Department of Genetics, Smurfit Institute
Trinity College Dublin
Dublin 2, Ireland
Tel: 353 1 6082319
Fax: 353 1 6798558
Email: swong@tcd.ie

CLASS: plasmid ends.
Location/Qualifiers
1..495
/organism="Kluveromyces delphensis"
/mol_type="genomic DNA"
/strain="CBS 2170"
/db_xref="taxon:51657"
/clone="K10625"
Library

ORIGIN
Query Match 13.8%; Score 66.8; DB 9; Length 495;
Best local similarity 56.5%; Pred. No. 7.1e-05;
Matches 166; Conservative 0; Mismatches 122; Indels 6; Gaps 2;

195 TTACAGCTGATATACGTAACTCATGTCGACCACTAATTTCTTCAACAGCGGTTCTG 254
82 TCAAAATTCATACAGTAAGTAAGGACCCATTAATTTCTCCAGCAGCTCTTCTGT 141

| | | | | |
|------------|--|------------------|---|----------------------------|
| Oy | | 255 | CTGTTTCCAAAGGACTCGGAATTAATGCCACATTATTCGTGGCAATTCATATTATATTC | 314 |
| Dd | | 142 | TGTTTGCCAAGGACACCAAAGATGCCCTATTAATCTTTGGCATTTACAACGGTATTC | 201 |
| Oy | | 315 | CCCTAATGGGATTAATGATGAGCCGCTTAA-----ATAAGCACCAGTAATATTCACCTT | 369 |
| Dd | | 202 | GTTAAGGTTAACCTTCATATTTGTGTAATTAATTTCCCATTTCATATTCAGTAGTAAT | 261 |
| Oy | | 370 | GTTGAGTTTTGTCTTTTGGCTTCTCTTAAGGCTTCATTTA-TCATAAGCAAGTTTTGT | 428 |
| Dd | | 262 | ACGACAGATATGAGGACTCTCAGTCAATAATTCATTAAAGCAATTCCAAAGAATTTTGGGT | 321 |
| Oy | | 429 | ATAATTCAAAATACCTTTCCTTTTCATGACTTGAAACCTCCAAATGATGAGGT | 482 |
| Dd | | 322 | ATATTCTAAATGCTGTTGCAGATCCATGTTTCAACCGCAAGATGATGT | 375 |
| RESULT 2 | | | | |
| LOCUS | CNS0064G | | | |
| DEFINITION | Drosophila melanogaster genome survey sequence T7 end of BAC # BACR13009 of RPCL-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence. | | | |
| ACCESSION | CNS0064G | | 1001 bp | DNA linear GSS 03-JUN-1999 |
| VERSION | AL062781 | | | |
| KEYWORDS | AL062781.1 GI:4943971 | | | |
| SOURCE | GSS. | | | |
| ORGANISM | Drosophila melanogaster (fruit fly) | | | |
| REFERENCE | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Prexygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. | | | |
| AUTHORS | Genoscope. | | | |
| TITLE | Direct Submission | | | |
| JOURNAL | Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr - Web : www.genoscope.cns.fr) | | | |
| COMMENT | Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCL-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain YZ; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm. | | | |
| FEATURES | | | | |
| source | location/Qualifiers | | | |
| | 1..1001 | | | |
| | /organism="Drosophila melanogaster" | | | |
| | /mol_type="genomic DNA" | | | |
| | /db_xref="taxon:7227" | | | |
| | /clone="BACR13009" | | | |
| | /clone_1ib="RPCL-98" | | | |
| | /note="end : T7" | | | |
| ORIGIN | | | | |
| | Query Match | 11.5%; | Score 55.6; | DB 10; Length 1001; |
| | Best Local Similarity | 38.3%; | Pred. No. 0.027; | |
| | Matches 157; | Conservative 42; | Mismatches 211; | Indels 0; Gaps 0 |
| Oy | 51 TATTTTACGTGATACCTTAACAAGTTGATATATGTTGTGTAACCTATTATTTAGACA | 110 | | |
| Dd | 54 TATTTTTCACATMYTTTTTYTWTTTTTYTTTTTTTTTTTTYTTMTYTYTTTTWTWTT | 113 | | |
| Oy | 111 GGATTTTAAACACCTTGAACATAAAGCTTAATAATAAATTTCTATCTTAAAG | 170 | | |

| | | | | |
|-----------------------|--|---|---|-----|
| Df | | 114 | ATTWTTTTCATTTTTTTVTVTAATTTTTMAAATTTAACTCAATAATTTTTTTTATNTC | 172 |
| OY | | 171 | GCACATTTTACGAGCGCTTAGGCCAATTTACAGCTGATATACGTAAAACTCATGTCCACT | 230 |
| Df | | 174 | ATTCACTTTTTTATTTCATTTTCATVTATTTTTTTWTNATAATTTTTTWCCTTTTTCTTWAVT | 233 |
| OY | | 231 | AAATCTCTTACACAGCGCTCTGCTCTTTCCAAGGACCTCCGAATATNGCACATTTAT | 290 |
| Df | | 234 | TTTTTTCTCTTCWCAATTTCTTCCTMTTTTTTTWTNAYTTTCTTCTTACAATTTAAATYT | 293 |
| OY | | 291 | CTGTGGCATTTCCAAATTTAATATTCCTTGATGGTATTTGANATGCGGTTTAAATAGTC | 350 |
| Df | | 294 | TTTTTTYTTTTTTTTTTTTTYVVVYVTVTWTTTTTTTTTTTTTTXYTTTTYYTTTTTTC | 353 |
| OY | | 351 | ACCAGTGAATCTTCACTTGTTGCAATTTGTCTTTGCTTCTCTAAAAGCTTCAATTT | 410 |
| Df | | 354 | YTTTTTTTTTTTTTTWTNATAATTTTAAATTTTYYTTTTTCTTTTTTTTWTNTTTTTTT | 413 |
| OY | | 411 | ATCTAAGAAGATTTTGATATATCAAATACCTTGCTTTTCCATGAC | 460 |
| Df | | 414 | TACTWTWCTATAYTTATATATTTTTTTWWAACITTTTTTTTATATTTWC | 463 |
| RESULT 3 CONS0039G | | CNS0039G | 1101 bp DNA linear GSS 03-JUN-1999 | |
| DEFINITION | | Drosophila melanogaster genome survey sequence TEI3 end of BAC # | | |
| LOCUS | | BACR08K10 of RPEC-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence. | | |
| ACCESSION | | AL063921 | | |
| VERSION | | GI:494178 | | |
| KEYWORDS | | GSS. | | |
| SOURCE | | Drosophila melanogaster (fruit fly) | | |
| ORGANISM | | Drosophila melanogaster | | |
| | | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; | | |
| | | Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; | | |
| | | Ephydroidea; Drosophilidae; Drosophila. | | |
| REFERENCE | | Genoscope. Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the iogenic strain Y2; cn bw ap, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm. | | |
| FEATURES | | location/Qualifiers | | |
| source | | 1..1101 | | |
| | | /organism="Drosophila melanogaster" | | |
| | | /mol_type="genomic DNA" | | |
| | | /db_xref="taxon:7227" | | |
| | | /clone_lib="BACR08K10" | | |
| | | /clone_1ib="RPCI-98" | | |
| | | /note="End : TEI3" | | |
| ORIGIN | | | | |
| Query Match | | 10.7%; Score 51.8; DB 10; Length 1101; | | |
| Best Local Similarity | | 15.6%; Pred.No. 0.2; | | |
| Matches | | 62; Conservative 181; Mismatches 154; Indels 0; Gaps 0 | | |
| OY | | 43 TATGAAATATATTTTACGTGTACTTTCAGAGTTCATATANGGTTGTGTGAATCTTATTT | 102 | |

[illegible]

| | | | | |
|------------|--|-----|--------|-----------------|
| CNS00EVL/c | 1101 bp | DNA | linear | GSS 04-JUN-1999 |
| LOCUS | | | | |
| DEFINITION | Drosophila melanogaster genome survey sequence T7 end of Bar: BACK29623 of RpCl-98 library from Drosophila melanogaster (fruit fly), genomic sequence. | | | |

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AL069706 GI:4949849
AL069706.1
GSS.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephyridae; Drosophilidae; Drosophila.

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage

COMMENT
BP 191 91006 EVRY cedex - FRANCE (E-mail: segre@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila genome project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila melanogaster* BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACpac Resource Center can be found at <http://bacpac.med.buffalo.edu/drosophila.bac.htm>.

```

FEATURES
source
location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR29823"
/clone_1fb="PCR-98"
/note="end : 77"

```

[illegible]

| RESULT 5 | CL104745 | LOCUS | DEFINITION |
|----------|------------|---------|---------------------------|
| | CL104745 | 1191 bp | DNA |
| | ISB1-33112 | | linear |
| | ISB1-33112 | | GSS 05-JAN-2004 |
| | | | genomic clone ISB1-43312, |
| | | | genomic library sequence. |

| | |
|-----------|---|
| ACCESSION | CL104745 |
| VERSION | CL104745.1 |
| KEYWORDS | GI:40598380 |
| SOURCE | GSS. |
| ORGANISM | Xenopus tropicalis (western clawed frog) |
| | Xenopus tropicalis |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |

REFERENCE
AUTHORS
TITLE

Amphibia; Barrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 1191)
Kremicki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
Mardis, E. and Wilson, R.
A physical map of the xenopus tropicalis genome

JOURNAL COMMENT
Unpublished (2003)
Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 75000 Std Error: 0.00

```

Seq primer: T7 TAATACGACTCACTATAGGG
Class: BAC ends
High quality sequence start: 36
High quality sequence stop: 397.
      location/qualifiers
FEATURES
    source
        1..1191

```

```

/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/db_xref="taxon:8364"
/clone="ISB1-43112"
/clone_1fb="ISB1"
/note="Vector: pBelorAC11; ISB-1 Xenopus tropicalis BAC

```

| ORIGIN | Library Segment 1" |
|-----------------------|--|
| Query Match | 9.9% ; Score 47.8; DB 10; Length 1191; |
| Best Local Similarity | 46.2% ; Pred. No. 1.6; |

2

[illegible]

Matches 192; Conservative 0; Mismatches 223; Indels 1; Gaps 1;

OY 43 TATGAAATATTTTACTGTGATCTTACAGTGTATGATGCTGTGTAATTT 102
|||
Db 606 TATTAATTTTATTTATTTATTTTAAAAATTTATATATATTTTATATTT 665
OY 103 ATTGAGAGTATTTTAAACACCTTAACTAAATTAATAAATAT-TTCTGA 161
666 TTAATATTTTATTTTATTTATTTATTTAAATTTTAAATTTTATTTATTT 725
OY 162 TCTTAAAGGACATATTAAGTGTAGGCAATTAAGCTGATAGCTGTAATCTAT 221
726 TTTTAAATTAATAATTAATTTAAATTTTAAATTTTAAATTTTAAATTT 785
OY 222 GTGGCAGTAAATCTTCTTACACCGCTTCTCTTCTTCAAGGAGCTCCGAATGCG 281
786 ATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAAATTTATTT 845
OY 282 ACTATTATCTGTGCAATTTCAATTTATTTCCCTATGGGATTTGATGCGCGTT 341
846 TTTATATTTTATTTTAAATTTTATTTATTTATTTTAAATTTTAAATTTT 905
Db 342 TAAATAGCAGGATTTGATCTTCTGAGTTTGTCTTGTCTTCTTAAAGT 401
906 TTTTATTTTATTTTATTTTATTTTATTTATTTATTTTATTTTATTTATTT 965
OY 402 CTTCATTTATCTTAAAGCAAGTTTGTATTAATTTCAAAATCTTGTCTTCTCAT 457
966 TTTTATTTTATTTTAAATTTTATTTTATTTAAATTTATTTATTTATTTAT 1021

RESULT 6
BH945927
LOCUS BH945927 705 bp DNA linear GSS 01-OCT-2002
DEFINITION chub8a02.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
sequence.
ACCESSION BH945927
VERSION BH945927.1 GI:23425987
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 705)
AUTHORS Delhaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T.,
Nash,W., Rabinowicz,P.D. and Wilson,R.K.
TITLE Whole genome shotgun reads from Brassica oleracea
JOURNAL Unpublished (2002)
COMMENT Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Plate: chub8 row: a column: 02
Seq primer: -21UpOT forward
Class: Shotgun
High quality sequence start: 26
High quality sequence stop: 551.
Location/Qualifiers
1..705
FEATURES
source
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/note="Vector: pOTw13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea T01000D3 buds provided by
Thomas Osborn at the University of Wisconsin. Genomic
DNA was provided by Pablo Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome
Sequencing Center."

ORIGIN

Query Match 9.8%; Score 47.6; DB 9; Length 705;
Best Local Similarity 45.5%; Pred.No.1.8;
Matches 170; Conservative 0; Mismatches 204; Indels 0; Gaps 0;

OY 53 TTTTACGTGATCTTACAGTGTATGATGCTGTGTAATTTATTTATTTGAGAG 112
|||
Db 225 TTTCACCTTGATGATGATGCAAGAAAAATTTAAAGTAAAGATTAACAAATATTTTATTTTA 284
OY 113 TATTTTAAACACCTTGAACCTTAATAATAATAATTTCTTATCTTTAAAGC 172
285 TTTTAAAGTATATTTTAAATTAATCTATTAATTAATTTTATTTATTTATTTA 344
OY 173 ACATTTACGCGCTAAGGCAATTAAGCTGATATCTGTTAACTCATGCGGCACATA 232
345 ATTCATCTCCAAAGTAAATCAATGAAAATTAATAATTAATAATAATCAATGAAGT 404
OY 223 ATTCCTTACACCGCTGCTCTCTTCTTCCAAAGGAGCTCCGAATATGCACTATTTATCT 292
405 TATTTTAAAGTATATTAATAATGTTCAATATATGACAAATTAATAGTTATATAT 464
OY 293 GTGGCATTTCCAAATTTATATCCCTATTTGGGATTTGATGCGCGTTTAAATAGTCAC 352
465 AAGATTTTAAACAATAATAATTTTATTAATAATTTTATTTATTTATTTCAATTTATA 524
OY 353 CGATTGAATCTTCACTTGTGAGTTTGTCTTTTGTCTTCTTAAAGTCTTCAATTTAT 412
525 TGATTTTCTTAAATTTTAAAGTTTACTTAATTAATTTGTAAATAATTTTAAACATTT 584
Db 413 CTAAAGCAAGTTT 426
585 GTAAACTTTGTTT 598

RESULT 7
CNS0106X/c
LOCUS CNS0106X/c 1101 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
BACN03K20 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL098595
VERSION AL098595.1 GI:5610206
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Peerygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage ;
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (BDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CBPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBeloBAC11.

FEATURES
source
Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone_lib="BACN03K20"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : T7"

ORIGIN

Query Match 9.7%; Score 47; DB 10; Length 1101;

Query Match

Best Local Similarity 9.5% ; Score 46.2; DB 1; Length 1161;

Matches 171; Conservative 0; Mismatches 227; Indels 2; Gaps 1;

Dy 51 TATTTTACTGTGTAACCTAACAAGTGATAATATGGTGTGAACCTATTATTTGAGA 110
||| |||
Db 756 TATTTTATGAGNGATAATATAATATATTTTTATTTTGAATATTTATTTTANN 815

Oy 111 GGATTTTAAACACCCTTAGAACCTAAAACCTTAATAAATATTCCTCATCCTTTAAAG 170
||| |||
Dy 816 GAATTAATTTTATPANTYAAAAANNTAAATANNNNNNTATATATTTTATNTNNNTAA 875

Oy 171 GCACATATPAACGTGGCTPPAGCAATTAACAGCTGATATPACTGPAAAACTCATGCGCAC 230
|||| |
Dy 876 GTAAATTTATATATTTATTTATTNNAAGTTTNGAATTTGATGATAPATATTTATATATAT 935

Oy 231 AAATCTCTCPAACGCCGTTCTGTCTCTTTCCAAAGGACTCCGAAPTGCACATATTAT 290
||| |||
Dy 936 TNATTTTATTAATAATGATTAATTCATTAATTTATTTTLAGNNNGTTTATATPATATTTGA 995

Oy 291 CTGGGCAATTTCCAATTTATATTCOCCTATTGGGATATTGAATGGCCCCGTTAAATAGTC 350
||| |||
Dy 996 GTAAG-ATTTTATATTTTTTTTNNCNTCNMTGGTAAATVATTTTTTTTTTAADATTTT 1053

Oy 351 ACCGATTAATCTCATCTGTGCGAGATTGTGTCTTGTCTCTCTCAAAGGCTTTCAATT 410
||| |||
Dy 1054 CTTNTTTTATTTNTNTNTATTTTATTTATTTGTMAATATATTPAANAATTTTATGATANTTA 1113

Oy 411 ATCTAAGCAAGTTTGNATPATYTCAAAATPACTTTGCTTT 450
||| |||
Dy 1114 ATTATATATGATGAATAAATGAATATGTTAANAATTTTATTTT 1153

RESULT 11
AJ928310/c
LOCUS
DEFINITION
VERSION
KEYWORDS
ACCESSION
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL COMMENT
FEATURES source

AJ928310 1079 bp mRNA linear EST 10-JUN-2005

AJ928310 Theileria annulata merozoite cDNA clone
tan017g02_gtk, mRNA sequence.

AJ928310
AJ928310.1 GI:67498696
EST.
Theileria annulata
Theileria annulata
Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae;
Theileria.
1 (bases 1 to 1079)

Pain,A., Renaud,H., Berriman,M., Murphy,L., Yeats,C.A., Weir,W., Kechornou,A., Ajalett,M., Bishop,R., Buchler,C., Cochet,M., Coulson,R.M.R., Cronin,A., de Villiers,E., Fraser,A., Foster,N., Gardner,M., Gobie,A., Griffiths-jones,S., Harris,D.E., Katzer,F., Lake,N., Lord,A., Maier,P., McKellar,S., Mooney,P., Morton,F., Nene,V., O'Neill,S., Price,C., Quail,M.A., Rabinowitsch,E., Rawlings,N.D., Kutcher,S., Saunders,J., Seeger,K., Shah,T., Squares,R., Squares,S., Tivey,A., Walker,A.R., Woodward,J., Döbelbauer,D.A.E., Langsley,G., Rajandream,M.-A., McKeever,D., Shiels,B., Taft,A., Barrettl,B. and Hall,N.

The genome of the host-cell transforming parasite Theileria annulata and a comparison wicth T. parva unpublished (2005)
Contact: Pain A
The Pathogen Sequencing Unit
The Wellcome Trust Sanger Institute
Genome Campus, CB10 1SR, UNITED KINGDOM
Merozoite cDNA library: Frank Katzner and Brian Shiels, Division of Veterinary Infection and Immunity, ICM, University of Glasgow, UK.
Location/Qualifiers
1..1079
/organism="Theileria annulata"
/mol_type="mRNA"
/isolate="Ankara (clone D7)"
/db_xref="taxon:5874"

| | | | | |
|-----------------------|--|---|-----------------|------------------------|
| Query Match | 9.5% | Score 46.2; | DB 1; | Length 1161; |
| Best Local Similarity | 42.8% | Prod. No. 3 8; | Mismatches 227; | Indels 2; Gaps 1; |
| Matches 171; | Conservative 0; | Mismatches 0; | Mismatches 227; | Indels 2; Gaps 1; |
| Dy | 51 | TATTTTACTGTGATACCTTAACAAGTGATATATATGGTGTGTACTTATTATTTGAGA | 110 | |
| | | | | |
| Db | 756 | TATTTTATGAGNGATATAATATATATTTTTATTTTGAATATATTTATTTTANN | 815 | |
| Oy | 111 | GGAATTTAAACACCCTTAGAATCAAAAACCTTAATAAATATTTCTCATCTTTAAAG | 170 | |
| | | | | |
| Db | 816 | GATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA | 875 | |
| Oy | 171 | GCACATATTAACGTGGCTPAAGGCAATTAACAGCTGATATACTGTAAAACTCATGTCCGCACT | 230 | |
| | | | | |
| Db | 876 | GTAATATTTATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT | 935 | |
| Oy | 231 | AAATCTTCTPAACACCGGTCCTGTCTCTTTCCAAGGACCTCCGAATATGCCACTATTTAT | 290 | |
| | | | | |
| Db | 936 | TNATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT | 995 | |
| Oy | 291 | CTGTGCAATTTCCAAATTTATATTCCTATTTGGGATTTTGAATGTGGCCGTTTAAATAGC | 350 | |
| | | | | |
| Db | 996 | GTAAG-ATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT | 1053 | |
| Oy | 351 | ACCATTAATTAATCTCACTGTGTGCAAGTTTGTCTTTGTCTTCTCAAAAGCTTTCAATTT | 410 | |
| | | | | |
| Db | 1054 | CTTTNTTTATTTNTTNTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTA | 1113 | |
| Oy | 411 | ATCTAAGCAAGTTTGTATTAATTCAAATATCTTTGCTTT | 450 | |
| | | | | |
| Db | 1114 | ATTAATATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT | 1153 | |
| RESULT 11 | | | | |
| AJ928310/c | | | | |
| LOCUS | AJ928310 | 1079 bp | mRNA | linear EST 10-JUN-2005 |
| DEFINITION | AJ928310 Theileria annulata merozoite Theileria annulata cDNA clone | | | |
| VERSION | AJ928310.1 GI:67498696 | | | |
| KEYWORDS | EST. | | | |
| SOURCE | Theileria annulata | | | |
| ORGANISM | Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae; | | | |
| REFERENCE | Theileria. | | | |
| AUTHORS | Pain,A., Renaud,H., Berriman,M., Murphy,L., Yeats,C.A., Weir,W., Kethornou,A., Ajalet,M., Bishop,R., Buchler,C., Cochet,M., Coulson,R.M.R., Cronin,A., de Villiers,E., Fraser,A., Foster,N., Gardner,M., Gobie,A., Griffiths-Jones,S., Harris,D.E., Katzer,F., Lake,N., Lord,A., Maser,P., McKellar,S., Mooney,P., Morton,F., Nene,V., O'Neill,S., Price,C., Quail,M.A., Rabinowitsch,E., Rawlings,N.D., Rutter,S., Saunders,D., Seeger,K., Shah,T., Squares,R., Squares,S., Tivey,A., Walker,A.R., Woodward,J., Dohelaer,D.A.E., Langsley,G., Rajandream,M.-A., McKeever,D., Shelds,B., Tait,A., Barrell,B. and Hall,N. | | | |
| TITLE | The genome of the host-cell transforming parasite Theileria annulata and a comparison with T. parva | | | |
| JOURNAL | Unpublished (2005) | | | |
| COMMENT | Contact: Pain A The Pathogen Sequencing Unit The Wellcome Trust Sanger Institute Genome Campus, CB10 1SR, UNITED KINGDOM Merozoite cDNA library: Frank Katzer and Brian Shiels, Division of Veterinary Infection and Immunity, ICM, University of Glasgow, UK. Location/Qualifiers 1..1079 /organism="Theileria annulata" /mol_type="mRNA" /isolate="Ankara (clone D7)" /db_xref="taxon:5874" | | | |
| FEATURES | source | | | |


```

/strain="Standard Poodle"
/db xref="taxon:9615"
/clone_lib="Dog library"
/note="Site 1: BstXI; Libraries were prepared from
peripheral blood"

```

ORIGIN

| Query Match | Score | DB | Length |
|--------------------------|-------|----------------|-------------------|
| Best Local Similarity | 59.2% | Pred. No. 6.4; | |
| Matches 77; Conservative | 0; | Mismatches 53; | Indels 0; Gaps 0; |

Dh
418 TTTTCTTTATATTTCGACCGAATTTTAACTGTATGAATGTAATAAAATTTTCCTTTTCTCCAAA 477

111 GGTATTTTAAACACACCTTAGAACTAAACTTAATAATAATATTTCTCTATCTTAAAG 170

Db 478 GCTATTTTGAACACAGTATTTCTAGAACATAATCTATATTTTATATTTACCGAAG 537

QY 171 GCACATATTA 180

Db 538 TACATATA 547

РЕСПУБЛИКА

| Accession | Length | Type | Source |
|-----------|--------|------------|--------|
| AV112242 | 215 bp | mRNA | linear |
| AV112243 | EST | 29-JUN-199 | |

clone 2610010G18, mRNA sequence.

VERSION AV112242.1 GI:5266322

SOURCE Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Eucarchontia; Claves; Rodentia;

REFERENCE 1 (pages 1 to 215)
Sciurognathi; Muroidea; Muridae; Mus.

AKAHIRA, S.: Akiyama, J.: Fukuda, S.: Fukunishi, Y.: Funayama, T.:
 FUJINOBU, E.: Ohtsuka, K.: Ozawa, T.: Kojima, H.: Itoh, M.: Hatawa, K.:
 FUJINOBU, E.: Ohtsuka, K.: Ozawa, T.: Kojima, H.: Itoh, M.: Hatawa, K.:

Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H.

Sugahara, Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y.,

Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

JOURNAL
Unpublished (1999)
COMMENT
Contact: Ch'ia O'ua

Genome Science Laboratory
RIKEN

Tel: 01-298-36-9145

Email: genome-res@rtc.riken.go.jp

trehalose and its application for the synthesis of full length cDNA

Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc Natl Acad Sci U S A 85(7):2455-2460 (1988))

For further details, please visit our web site (<http://genome.rcc.rkenl.go.jp/>)

| FEATURES | LOCATION/Qualifiers |
|----------|---------------------|
| SOURCE | 1. .215 |

```

/mol type="mRNA"

```

```
/db_xref="taxon:10090"
```

```
/sex="mixed"
```

```
/clone_lib="Mus musculus C57BL/6J 10-day embryo"
```


Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 28, 2005, 09:00:35 ; Search time 157 Seconds

(without alignments)
5491.194 Million cell updates/sec

Title: US-10-600-230-1

Sequence score: 485 1 gcgcgcggaaggaagaagt.....cctccaatgatgaggtacc 485

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/prodata/1/ina/1/COMB.seq:*
2: /cgn2_6/prodata/1/ina/5/COMB.seq:*
3: /cgn2_6/prodata/1/ina/6/COMB.seq:*
4: /cgn2_6/prodata/1/ina/6/COMB.seq:*
5: /cgn2_6/prodata/1/ina/6/COMB.seq:*
6: /cgn2_6/prodata/1/ina/6/COMB.seq:*
7: /cgn2_6/prodata/1/ina/6/COMB.seq:*
8: /cgn2_6/prodata/1/ina/6/COMB.seq:*
9: /cgn2_6/prodata/1/ina/6/COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|---------------------|-------------------|
| 1 | 43.8 | 9.0 | 23669 | US-09-949-016-15296 | Sequence 15296, A |
| 2 | 43.8 | 9.0 | 23672 | US-09-949-016-12575 | Sequence 12575, A |
| 3 | 43.8 | 9.0 | 23672 | US-09-949-016-14941 | Sequence 14941, A |
| 4 | 42 | 8.7 | 832 | US-09-621-975-2813 | Sequence 2813, Ap |
| 5 | 41.6 | 8.6 | 1358 | US-09-270-767-29720 | Sequence 29720, A |
| 6 | 41.6 | 8.6 | 1358 | US-09-270-767-29720 | Sequence 29720, A |
| 7 | 41.6 | 8.6 | 1358 | US-09-270-767-29720 | Sequence 29720, A |
| 8 | 41 | 8.5 | 705 | US-10-030-808-1 | Sequence 14, Appl |
| 9 | 40.6 | 8.4 | 111235 | US-09-949-016-15328 | Sequence 15328, A |
| 10 | 40.2 | 8.3 | 111701 | US-09-949-016-13314 | Sequence 13314, A |
| 11 | 39.8 | 8.2 | 640681 | US-09-790-988-1 | Sequence 1324, A |
| 12 | 39.6 | 8.2 | 786431 | US-09-751-389-3 | Sequence 1, Appl |
| 13 | 39.2 | 8.1 | 981 | US-09-248-796A-1889 | Sequence 3, Appl |
| 14 | 39 | 8.0 | 564 | US-09-248-796A-1889 | Sequence 1889, Ap |
| 15 | 39 | 8.0 | 174170 | US-09-949-016-14810 | Sequence 14810, A |
| 16 | 39 | 8.0 | 174170 | US-09-949-016-14810 | Sequence 14810, A |
| 17 | 39 | 8.0 | 174170 | US-09-949-016-14810 | Sequence 14810, A |
| 18 | 39 | 8.0 | 174318 | US-09-949-016-14812 | Sequence 14812, A |
| 19 | 39 | 8.0 | 174318 | US-09-949-016-14813 | Sequence 14813, A |
| 20 | 38.4 | 7.9 | 1215 | US-08-642-274D-35 | Sequence 35, Appl |
| 21 | 38.4 | 7.9 | 1215 | US-08-952-014C-35 | Sequence 35, Appl |
| 22 | 38.4 | 7.9 | 146401 | US-09-949-016-16151 | Sequence 16151, A |
| 23 | 38.2 | 7.9 | 146428 | US-09-949-016-12620 | Sequence 12620, A |
| 24 | 38.2 | 7.9 | 146428 | US-09-949-016-12081 | Sequence 12081, A |

| | | | | | |
|----|------|-----|---------|---------------------|-------------------|
| 25 | 37.8 | 7.8 | 186959 | US-09-949-016-13125 | Sequence 13125, A |
| 26 | 37.6 | 7.8 | 237863 | US-09-949-016-13404 | Sequence 13404, A |
| 27 | 37.4 | 7.7 | 1316 | US-09-270-767-2481 | Sequence 2481, Ap |
| 28 | 37.4 | 7.7 | 1316 | US-09-270-767-17763 | Sequence 17763, A |
| 29 | 37.4 | 7.7 | 1856 | US-09-704-725-4 | Sequence 4, Appl |
| 30 | 37.4 | 7.7 | 199471 | US-09-949-016-14083 | Sequence 14083, A |
| 31 | 37.4 | 7.7 | 1664976 | US-08-916-421B-1 | Sequence 1, Appl |
| 32 | 37.4 | 7.7 | 1664976 | US-09-692-570-1 | Sequence 1, Appl |
| 33 | 37.2 | 7.7 | 94830 | US-09-949-016-12414 | Sequence 12414, A |
| 34 | 37.2 | 7.7 | 94830 | US-09-949-016-16336 | Sequence 16336, A |
| 35 | 37.2 | 7.7 | 144922 | US-09-949-016-15890 | Sequence 15890, A |
| 36 | 37 | 7.6 | 601 | US-09-949-016-80944 | Sequence 80944, A |
| 37 | 37 | 7.6 | 1141 | US-09-806-708B-22 | Sequence 22, Appl |
| 38 | 37 | 7.6 | 44353 | US-09-949-016-15302 | Sequence 15302, A |
| 39 | 37 | 7.6 | 50000 | US-09-662-254B-23 | Sequence 23, Appl |
| 40 | 36.8 | 7.6 | 18651 | US-09-949-002-592 | Sequence 592, App |
| 41 | 36.8 | 7.6 | 18682 | US-09-949-002-786 | Sequence 786, App |
| 42 | 36.8 | 7.6 | 20674 | US-09-641-638-651 | Sequence 651, App |
| 43 | 36.8 | 7.6 | 20674 | US-10-170-097-651 | Sequence 651, App |
| 44 | 36.8 | 7.6 | 35646 | US-09-949-016-11896 | Sequence 11896, A |
| 45 | 36.8 | 7.6 | 35647 | US-09-949-016-13360 | Sequence 13360, A |

ALIGNMENTS

```
RESULT 1
US-09-949-016-15296/c
Sequence 15296, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C1001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15296
LENGTH: 23669
TYPE: DNA
ORGANISM: Human
US-09-949-016-15296
Query Match
Best Local Similarity 9.0%; Score 43.8; DB 3; Length 23669;
Matches 78; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
43 TATGAATAATTTTACTCTGATCTTCAAGTGTATATGCTGTGTAATTT 102
9808 TATATAAAGATATATATAAGATATATAAGATATATATAAGATATATAT 9749
103 ATTGAAGAGTATTTTAAACACCTTAGAAGTAACTTAATTAATTAATTTCTAT 162
9748 AATATTAAGATATATATATATATCTTAATATATATATTAAGATATATATAG 9689
163 CTTTAAGGACATA 177
9688 AATATTAAGATATATA 9674
RESULT 2
US-09-949-016-12575/c
Sequence 12575, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
```


CURRENT FILING DATE: 1999-02-12

471 CATTCCCTTATTCCTTCTCC 445

```

RESULT 15
US-09-949-016-14810/c
: Sequence 14810, Application US/09949016
: Patent No. 6812339
: GENERAL INFORMATION:
: APPLICANT: VENTER, J. Craig et al.
: TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
: TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: CLO001307
: CURRENT APPLICATION NUMBER: US/09/949,016
: PRIOR FILING DATE: 2000-04-14
: PRIOR APPLICATION NUMBER: 60/241,755
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 60/237,768
: PRIOR FILING DATE: 2000-10-03
: PRIOR APPLICATION NUMBER: 60/231,498
: PRIOR FILING DATE: 2000-09-08
: NUMBER OF SEQ ID NOS: 207012
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 14810
: LENGTH: 174170
: TYPE: DNA
: ORGANISM: Human
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)_(174170)
: OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14810

```


Db 241 AACAGCGTTCGTCTCTTCCAGGAGACTCCGAATATGCCACTATTATCTGTGGCAAT 300
Qy 301 TCCAATTATATATCCCTATATGGGATATTTGATGTGGCCGTTTAAATAGTACCGATGGA 360
Db 301 TCCAATTATATATCCCTATATGGGATATTTGATGTGGCCGTTTAAATAGTACCGATGGA 360
Qy 361 TCTTCACTTGTTCGAGTTTGTCTTTGCTTCTCTTAAAGTCTTCAATTATCTTAAAGA 420
Db 361 TCTTCACTTGTTCGAGTTTGTCTTTGCTTCTCTTAAAGTCTTCAATTATCTTAAAGA 420
Qy 421 AGTTTGTATATTAATCAAAATACCTTCTTCTCCATGACTTGAACCTTCAATGATGAG 480
Db 421 AGTTTGTATATTAATCAAAATACCTTCTTCTCCATGACTTGAACCTTCAATGATGAG 480
Qy 481 GTACC 485
Db 481 GTACC 485

RESULT 2

US-10-600-230-81
; Sequence 81, Application US/10600230
; Publication No. US20040092020A1
; GENERAL INFORMATION:
; APPLICANT: Wilkinson, Jack
; APPLICANT: McBride, Kevin
; APPLICANT: Bertalan, Sean
; TITLE OF INVENTION: GENETIC CONSTRUCTS HAVING HETEROLOGOUS
; TITLE OF INVENTION: 3' POLYADEMNYLATION SIGNAL SEQUENCE MOTIFS THAT FUNCTION IN
; FILE REFERENCE: 0325.210
; CURRENT APPLICATION NUMBER: US/10/600,230
; CURRENT FILING DATE: 2003-06-20
; PRIOR APPLICATION NUMBER: 60/390,529
; PRIOR FILING DATE: 2002-06-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 81
; LENGTH: 4046
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-10-600-230-81

Query Match 97.9%; Score 474.8; DB 7; Length 4046;
Best Local Similarity 99.6%; Pred. No. 3.8e-101;
Matches 476; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 6 CGGAAGAGGAAAGTGCCTCGTTGCTAGACATGATGAAATATTTTACTGTGAT 65
Db 3546 CGGAAGAGGAAAGTGCCTCGTTGCTAGACATGATGAAATATTTTACTGTGAT 3605
Qy 66 ACTTACAAGTTGATATATATGTTGTGTACTTATTTATTTGAGAGTATTTTAAACAC 125
Db 3606 ACTTACAAGTTGATATATATGTTGTGTACTTATTTATTTGAGAGTATTTTAAACAC 3665
Qy 126 CTTAGAAGCTTAAAGCTTATATTAATATTTCTTCTTAAAGGACATTTAGGTGG 185
Db 3666 CTTAGAAGCTTAAAGCTTATATTAATATTTCTTCTTAAAGGACATTTAGGTGG 3725
Qy 186 CTAAAGCAATTAAGCTGATATCTGTAAAGTCAATGCGGCACTAAATCTTCTAACG 245
Db 3726 CTAAAGCAATTAAGCTGATATCTGTAAAGTCAATGCGGCACTAAATCTTCTAACG 3785
Qy 246 GCGTCTGTCTCTTCCAAAGGACTCCGAATATGCCATTTATCTGTGGCATTTCCAA 305
Db 3786 GCGTCTGTCTCTTCCAAAGGACTCCGAATATGCCATTTATCTGTGGCATTTCCAA 3845
Qy 306 TTTATATATCCCTATATGGGATTTGATGTGGCCGTTTAAATAGTACCGATGATCTTC 365
Db 3846 TTTATATATCCCTATATGGGATTTGATGTGGCCGTTTAAATAGTACCGATGATCTTC 3905
Qy 366 ACTGTGTGAGTTTGTCTTTGCTTCTTAAAGTCTTCAATTTATCTTAAAGCAAGTTT 425

Db 3906 ACTTGTGAGTTTGTCTTTTGTCTCTTAAAGTCTTCAATTTATCTTAAAGCAAGTTT 3965
Qy 426 TGTATATTAATCAAAATACCTTCTTCTTCTCCATGACTTGAACCTTCAATGATGAGT 483
Db 3966 TGTATATTAATCAAAATACCTTCTTCTTCTCCATGACTTGAACCTTCAATGATGATG 4023

RESULT 3

US-10-027-632-114994/c
; Sequence 114994, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 114994
; LENGTH: 3254
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-114994

Query Match 9.7%; Score 47; DB 5; Length 3254;
Best Local Similarity 55.1%; Pred. No. 1.1; Indels 0; Gaps 0;
Matches 92; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
Qy 289 ATCTGTGCAATTCGAATTAATATCCCTATATGGGATATTTGATGTGGCCGTTTAAATAG 348
Db 1492 ATCTTAATTAATGAATATGATTTTCTTCTTGAATTTTAAATTTGGCTCTTTTAAAG 1433
Qy 349 TCACGATTTGAATCTTCACTTGTGAGTTTGTCTTGTCTTAAAGTCTTCAAT 408
Db 1432 TCTTATATTTCTATGCTGAAATCTTCATCTTTTCTTAAATCTTAAATGTTATGAT 1373
Qy 409 TTTATCAAGCAAGTTTGTATATTAATCAAAATACCTTCTTCTTCTCC 455
Db 1372 TTTATTAAGATCTTGTATGTTAATCTTATATGTTGTTGCTTC 1326
RESULT 4
US-10-027-632-114994/c
; Sequence 114994, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483

```

; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 114994
; LENGTH: 3254
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-114994
```

```
Query Match
Best Local Similarity 9.7%; Score 47; DB 6; Length 3254;
Matches 92; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
```

```
QY 289 ATCTGGGCAATTCATTATATATCCCTATTGGGATTTGATGGCCGTTAAATAG 348
DB 1492 ATCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1433
QY 349 TCACCGAATGAATCTTCACTTGTTCGAGTTTGTCTTCTCTAAAGTCTTCAAT 408
DB 1432 TCTTCAATTTCTAGCGAATTCCTCATCTTTTCTTAAATGTTAATGAT 1373
QY 409 TTATCTAAGCAAGTTTGTATTAATCAAAATCTTGTCTTCTC 455
DB 1372 TTATTTAAAGATCTGTATGTATTAATCTATATGTATGTCTTC 1326
```

RESULT 5

```
US-10-027-632-179264/c
; Sequence 179264, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 179264
; LENGTH: 1223197
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1223197)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-179264
```

```
Query Match
Best Local Similarity 9.6%; Score 46.4; DB 5; Length 1223197;
Matches 28.0%; Pred. No. 13;
```

Matches 104; Conservative 86; Mismatches 182; Indels 0; Gaps 0;

```
QY 78 AATATAGTGTGTGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 137
DB 1195986 ATMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMT 1195927
QY 138 ACTTAATAATAATTAATTTCTATCTTAATTAAGCAATATTAACGTGCTAAGCAATTA 197
DB 1195926 TMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMT 1195867
QY 198 CAGCTAATATACGTAAACATCATGCGCACTAAATCTTCTAACAAGCGTTCGTCTC 257
DB 1195866 TMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMT 1195807
QY 258 TTTCGAAGGACTCGGAATGCGCACTAATTAATCTGTGGCAATTCATTAATTTCC 317
DB 1195806 TMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMT 1195747
QY 318 TATGGGATTTGAATGTCGCCGTTAAATAGTCAACCGATGATCTTCACTGTTCGACT 377
DB 1195746 TMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMT 1195687
QY 378 TTGTCTTTGCTCTCTTAAGGCTTCAATTAATTAATTAATTAATTAATTAATTAAT 437
DB 1195686 TMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMT 1195627
QY 438 AATACCTTGCTT 449
DB 1195626 GCTGTGTGCTT 1195615
```

RESULT 6

```
US-10-027-632-179264/c
; Sequence 179264, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 179264
; LENGTH: 1223197
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1223197)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-179264
```

Query Match
Best Local Similarity 9.6%; Score 46.4; DB 6; Length 1223197;
Matches 104; Conservative 86; Mismatches 182; Indels 0; Gaps 0;

QY 78 AATATAGTGTGTGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 137

```

Db      908  TTATTTTTTTTTTGAAGATTGTTTTATTAATAATTTTTTGTTTTTTTTTTAGTTT    967
QY      271  CCGAATATGCACATATTATCTGTGGCAATTCCAAATTATATCCCATTTGGGTATTG    330
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      968  TTTTATATAAATTTTTTGTAGTATTTTTTTGTGTACGTTTTAAAAAGTTGTATG    1027
QY      331  ATGTGGCGGTTAAATATAGTCACGATTAATCTTCATGTTCGAGTTTTGCTTTGCT    390
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1028  TTAAGTTTGTATATGATTTTTTTTTTTTTTTTTTTTTTTTAAATTTTTTTTTTTTT    1087
QY      391  TCTCTAAAGTCTTCATATTATCTAAAGCAAGTTTGTATATCCAAATATCTTGCTT    450
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1088  TTTTATATGTTTTTTTTTAAATTTTTTTTTTAAATATTTTATAGTTTGTAGATTTTATTT    1147
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      451  TCT 453
      ||| |||
Db      1148  TAT 1150

RESULT 8
US-10-221-613-36
; Sequence 36, Application US/10221613
; Publication No. US20040029123A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander

```

```

? TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
? FILE REFERENCE: 5013.1004
? CURRENT APPLICATION NUMBER: US/10/221,613
? CURRENT FILING DATE: 2002-09-13
? PRIOR APPLICATION NUMBER: PCT/EP01/02945
? DE 10013847.00
? DE 10019058.8
? DE 10019173.8
? DE 10032528.7
? DE 10043826.1
? PRIOR FILING DATE: 2001-03-15
? 2000-03-15
? 2000-04-06
? 2000-04-07
? 2000-06-30
? 2000-09-01
? NUMBER OF SEQ ID NOS: 428
? SEQ ID NO 36
? LENGTH: 6092
? TYPE: DNA
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-613-36

Query Match          9.5%;      Score 46.2;  DB 7;   Length 6092;
Best Local Similarity 46.8%;      Pred. No. 2.1;

```

```

QY      62  TGAATACTACAGGTGATATATGTTGTGTGAACCTATATTTATTTAGAGAGTATTTAAAC  121
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      4944  TGTATTTTAAAGTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT  5001
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      122  ACACCTTAGACATAAACAATTAAATAAATAATTTCTATCTTTAAAGGACATATTAAC  181
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      5004  ATTTTGTTTTAAATAAATTTGGAAATAATTTAGTTTATATATTAATTAAGTTTGAATAT  5066
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      182  GTGGCTAAGGCATTACAGCTGATATACGTGAAAACATGTCGGCACATAATCTTCTA  241
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      5064  AAATTGATATATATATAAATAATTTGATATTTTTTATTAATAATTTTGTTTTAAAAATTTTAA  5122
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      242  AACAGGCTGTGCTCTTTCCAAAGGACCTCCGAATATGCGCATATTTATCTGTGGCATTT  301
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      5124  TGTGTATTTTATATATTTATATATGTATATAAATAATGAA--TAAATATTTGTGTGAAT  5186
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

| | | | | |
|----|------|--------------------|---|------|
| Qy | 302 | CCAAATTATATATTC | CCCGATATGGGATATTTGATGTGGCGCGTTAAATAGTCACCGATTGAAT | 361 |
| | | | | |
| Db | 5181 | ATAGTTTAAGTTTAAAG | TTTAAAGTTTAAAGATTATATTTATTTATATTTAGTGTGTTTATATA | 5240 |
| | | | | |
| Qy | 362 | CTTCACTTGTTCGAGTTT | TGCTCTTTCCTCTCTAAAGGCTTCAATTTATCTAAAGCAA | 421 |
| | | | | |
| Db | 5241 | TATAGATTATTTATATAG | TATATTTATTTTATATATTTGATTTTATTAAGTAA | 5300 |
| | | | | |
| Qy | 422 | GTTTGTATATCTCAAA | TACTCTTGCT | 448 |
| | | | | |
| Db | 5301 | AGAAATTAAATATTAAAG | AGATTGAT | 5327 |
| | | | | |

RESULT 9

US-10-433-793-28

Sequence 28, Application US/10433793

Publication No. US20040142334A1

GENERAL INFORMATION:

APPLICANT: Epigenomics AG

TITLE OF INVENTION: Diagnose von mlt Angiogenese assoziierten Krankheiten

FILE REFERENCE:

CURRENT APPLICATION NUMBER: US/10/433,793

CURRENT FILING DATE: 2003-06-06

NUMBER OF SEQ ID NOS: 212

SEQ ID NO 28

LENGTH: 37515

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-433-793-28

| | | | | |
|-----------------------|-------|----------------|-------|----------------|
| Query Match | 9.3% | Score 45; | DB 7; | Length 37515; |
| Best Local Similarity | 53.0% | Pred. No. 7.8; | | |
| Matches | 96; | Conservative | 0; | Mismatches 85; |
| | | | | Indels 0; |
| | | | | Gaps 0; |

| | | | |
|----|-------|--|-------|
| Qy | 43 | TATGAAAAATATTTTACTGTGATCACTTACAAAGTTGATATATGTTGTGCTAACTTATTT | 102 |
| Db | 14019 | TATATATTTTTTTTATTTATTTAGATTAAATTTTATTTATATTTTTTGTATGAGATGTAATT | 14078 |
| Qy | 103 | ATTGAGAGGATTTTAAACACCTTGAACATAACTTAATAATAATTTTCCTAT | 162 |
| Db | 14079 | GTTTGTAGATTAATAAAAAATTTTTTAAATTTTAAATTTATATATATTTGTGTT | 14138 |
| Qy | 163 | CTTTAAAGGACATATTACGGGTAAAGCAATTAACAGTGATATACGTGAAAACTCAG | 222 |
| Db | 14139 | AATTTTTTGTAAATTTTCGGATGTTTAAGAAATTTGAATGAGTAGGTAAATGATATTAAG | 14198 |

| | | | |
|----|-------|---|-------|
| Qy | 223 | T | 223 |
| Db | 14199 | T | 14199 |

```

RESULT 10
US-10-311-455-309
: Sequence 309. Application US/10311455
: Publication No. US20030143606A1
: GENERAL INFORMATION:
: APPLICANT: ODEK, Alexander
: APPLICANT: PIEPERROCK, Christian
: APPLICANT: BERLIN, Kurt
: TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
: TITLE OF INVENTION: Cytosine methylation
: FILE REFERENCE: 5013.1014
: CURRENT APPLICATION NUMBER: US/10/311,455
: CURRENT FILING DATE: 2002-12-16
: PRIOR APPLICATION NUMBER: PCT/EP01/07537
: PRIOR FILING DATE: 2001-07-02
: PRIOR APPLICATION NUMBER: DE 10032529.7
: *PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: DE 10043826.1
: PRIOR FILING DATE: 2000-09-01
: NUMBER OF SEQ ID NOS: 2424

```

```

; SEQ ID NO 309
; LENGTH: 7201
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
DS-10-311-455-309

```

| | | | | |
|-----------------------|-----------------|----------------|-----------|--------------|
| Query Match | 9.2% | Score 44.8; | DB 6; | Length 7201; |
| Best Local Similarity | 53.4%; | Pred. No. 4.8; | | |
| Matches 94; | Conservative 0; | Mismatches 82; | Indels 0; | Gaps 0; |

Oy 14 GGAAGAGCACCCCTCGTGGGTGAGACAGTGTGAATAATATTTTAACGTGTACTTAACA 73
 Db 3407 GAAAAGACATTAATTAAGTTTTATTTAAGGCAATTTATTTTTTTTAATTATTTATTCGAATG 3466
 Oy 74 GTTGATATATGTTGTGTGTAACCTAATTATTTAGAGGTAATTTTAACACACCTTAGAAC 133
 Db 3467 GTTTTTTAATCTTTTTTTTTTTTATTTTGTTTTTTTTATTAAGTTTAACTTTTGTGGTTT 3522
 Oy 134 TAAACTTAATTAATTAATATTTCTCTATCTTTAAGGACATATTAAGTGGCTAA 189
 Db 3527 TAAATGCAATTTAATATATTTTTTATTTTTTTTATTAATTAATTTTAAGTGAATTA 3582

RESULT 11
US-10-363

US-10-363-345A-1843
; Sequence 1843, Application US/10363345A
; Publication No. US20040234960A1

```

; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Method for determining the degree of methylation of defined
; TITLE OF INVENTION: cytosines in genomic DNA in the sequence context of 5'-CpG-3'
; FILE REFERENCE: E01/1227
; CURRENT APPLICATION NUMBER: US/10/363,345A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 1843
; LENGTH: 517
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: CpG-island No: 1843
US-10-363-345A-1843

```

| | | | | |
|---------------------------|--------|-----------------|-----------|-------------|
| Query Match | 9.2% | Score 44.4; | DB 8; | Length 517; |
| Best Local Similarity | 44.6%; | Pred. No. 2.3; | | |
| Matches 174; Conservative | 0; | Mismatches 216; | Indels 0; | Gaps 0 |

| | | | |
|----|-----|--|-----|
| QY | 44 | ATGAAAATATTTTACTGTAATCTTACAAGTGTAAATGGTGTGAGTAATATTTA | 103 |
| Db | 63 | ATTATATATTTTCGTAATTTTATATTTAGATATTTATATTTTATTTATTTTCGT | 122 |
| QY | 104 | TTTGAGAGTAATTTTAAACAACCTTAGAATPAAACCTTAATTAATTAATATTTCTATC | 163 |
| Db | 123 | TTTTCGTATTTTTTTTTTATATTTTTATTAATAATAATTTGTAATCGTCGTTTTTTTTTC | 182 |
| QY | 164 | TTTAAAGCACATATTTACGTGGCTAAGGCAATTAACGCTGATATATCTGTAATACATGT | 223 |
| Db | 183 | GTAATATATATATATTTTCGGTCGAAGTATATCGTTATATACGTTTTTTTTTGTTGCT | 242 |
| QY | 224 | CGCACATAATTTCTTCAACACGGCTCTGTCTTTTCCAAAGGAAGCTCCGAATATAGCAC | 283 |
| Db | 243 | ATGTATTTAAATTTTTTTTTTCGTATATTTTCGTATTTATTTATTTTTTTTTTTCGC | 302 |
| QY | 284 | TATTTATCGTGGCAATTTCAATATATATATCCCATTTGGGATTTGATGTGGCGTTTA | 343 |
| Db | 303 | GATTTATTTTTTTTATGTTTTTTTTTTTTTTTACGATTTACGATATATTTTTTTATATAG | 362 |
| QY | 344 | AATAGTCCAGATGAAATCTTACCTGTTCGAGTTTTGCTCTTTTGCTCTCTAAAGGTCT | 403 |

Db 363 TATTATATATATTTTGGTTTAAATTTTTCGTTTTTATTTTATATATATATAT 422
QY 404 TCAATTTATCTAAGCAAGTTTGTATAT 433
Db 423 CGTTTTTTTTACGTTATTTTCGTTTAT 452

RESULT 12

US-10-363-345A-1844/c
; Sequence 1844, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Method for determining the degree of methylation of defined
; FILE REFERENCE: E01/1227
; CURRENT APPLICATION NUMBER: US/10/363,345A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 1844
; LENGTH: 517
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: Cpg-Island No: 1844
US-10-363-345A-1844

Query Match 9.2%; Score 44.4; DB 8; Length 517;
Best Local Similarity 44.6%; Pred. No. 2.3; Matches 174; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

QY 44 ATGAAATATTTTCTGATCTACAGTTGATATGCTGCGTAACCTATTTA 103
Db 455 ATTTATATTTTTCGTTAAATTTTATATATATTTTATTTTATTTTATTTTTCGTT 396
QY 104 TTTGAGAGTATTTTAAACACCTTAGAATAAATAATTAATTTCTATC 163
Db 395 TTTTGGTATTTTATTTTATTTATATATATATATATATATATATATATATATATAT 336
QY 164 TTTAAAGCAGATTTACGTGCTAAGCAATTAAGCTGATTAATCTGTAAGCTCATGT 223
Db 335 GTAT 276
QY 224 CGGCATAATCTTCTAAGAGCGTTCCTCTTCCAGGAGCTCCGAATATGCCAC 283
Db 275 ATGTAT 216
QY 284 TATTATCTGTGACATTTCAATTTATATCCCTATATGAGTATGAGCGCTTAA 343
Db 215 GATTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAG 156
QY 344 AATAGTACCGATGAACTTCTCTGTTGAGTTTGTCTTCTCTTAAGGCTCT 403
Db 155 TATTATATATTTTTCGTTTATATTTTTCGTTTATATTTTATATATATATATATAT 96
QY 404 TCAATTTATCTAAGCAAGTTTGTATAT 433
Db 95 CGTTTTTTTTTACGTTATTTTCGTTTAT 66

RESULT 13

US-10-363-483A-1843
; Sequence 1843, Application US/10363483A
; Publication No. US20050064401A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain

; TITLE OF INVENTION: illnesses
; FILE REFERENCE: 82011
; CURRENT APPLICATION NUMBER: US/10/363,483A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 1843
; LENGTH: 517
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: Cpg-Island No: 1843
US-10-363-483A-1843

Query Match 9.2%; Score 44.4; DB 9; Length 517;
Best Local Similarity 44.6%; Pred. No. 2.3; Matches 174; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

QY 44 ATGAAATATTTTCTGATCTACAGTTGATATGCTGCGTAACCTATTTA 103
Db 63 ATTTATATTTTTCGTTAAATTTTATATATATTTTATTTTATTTTATTTTATTTTTCGTT 122
QY 104 TTTGAGAGTATTTTAAACACCTTAGAATAAATAATTAATTTCTATC 163
Db 123 TTTTGGTATTTTATTTTATTTTATATATATATATATATATATATATATATATATAT 182
QY 164 TTTAAAGCAGATTTACGTGCTAAGCAATTAAGCTGATTAATCTGTAAGCTCATGT 223
Db 183 GTAT 242
QY 224 CGGCATAATCTTCTAAGAGCGTTCCTCTTCCAGGAGCTCCGAATATGCCAC 283
Db 243 ATGTAT 302
QY 284 TATTATCTGTGACATTTCAATTTATATCCCTATATGAGTATGAGCGCTTAA 343
Db 303 GATTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAG 362
QY 344 AATAGTACCGATGAACTTCTCTGTTGAGTTTGTCTTCTCTTAAGGCTCT 403
Db 363 TATTATATATTTTTCGTTTAAATTTTTCGTTTATATTTTATATATATATATATAT 422
QY 404 TCAATTTATCTAAGCAAGTTTGTATAT 433
Db 423 CGTTTTTTTTTACGTTATTTTCGTTTAT 452

RESULT 14

US-10-363-483A-1844/c
; Sequence 1844, Application US/10363483A
; Publication No. US20050064401A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
; FILE REFERENCE: 82011
; CURRENT APPLICATION NUMBER: US/10/363,483A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 1844
; LENGTH: 517
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: Cpg-Island No: 1844
US-10-363-483A-1844

Query Match 9.2%; Score 44.4; DB 9; Length 517;
Best Local Similarity 44.6%; Pred. No. 2.3; Matches 174; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

| | | | |
|----|-----|--|-----|
| OY | 4 | ATGAAATATTTTACGGATACCTACAGGTGAAATAGTGTGTGAACTATT | 103 |
| Db | 455 | ATTATATTTTTTGGTAAATTTTATTTAGATATTTTATTTATTTTAAATTTTCGTT | 396 |
| OY | 104 | TTTGAGAGATTTTAAACAACCTTGAACTAAACTTAATTAATAATATTTCTATTC | 163 |
| Db | 395 | TTTTTCGTATTTTTTTTATTTATTTTAAATATATATATTCGTATACGTCGTTTTTTTTTC | 358 |
| OY | 164 | TTTAAAGGCACATTTACGTGCTAAGSCAATTAACGCTATATACGTAAACTCATGT | 223 |
| Db | 335 | GTATATATTATTTATTTCCGCTGGAAGTATTCGTTTAAATCGTTTTTTTTTGTTGGT | 276 |
| OY | 224 | CGCAGTAATTTCTCTAACACGCGTCTGTCTCTTTCCAAAGCCTCCGAATATGCAAC | 285 |
| Db | 275 | ATGTATTTAAATTTTTTTTTTTCGTTATTTTGTATTTATTTTATTTTTTTTTTTCGC | 216 |
| OY | 284 | TATTTATCTGSCATTTCCAAATTTATTTCCCTATTTGGGATTTGAATGGCCGTTTA | 343 |
| Db | 215 | GATTTATTTTTTTTAAGTTTTTTTTTTTTTTTAAAGTATTTACGATATATTTTTTAAATACG | 156 |
| OY | 344 | AATATGACCGATTTGAATCTCACTTTGTCAGATTTTGCTTTTGCTTCTGAAAGTCT | 403 |
| Db | 155 | TATATATATTTTTCGTTTTTAAATTTTTTTTCGTTTTTAAATTTTAAATATATTTAT | 96 |
| OY | 404 | TCAATTAATCTAAGCAGATTTGTTAAT | 433 |
| Db | 95 | CGTTTTTTTTTACGTTATTTTTCGTTAAT | 66 |

RESULT 15
US-09-925-065A-53227

```

Sequence 53227, Application US/09925065A
Publication No. US20050228172A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 53227
LENGTH: 1803
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-53227

```

| | | | | |
|---------------------------|--------|-----------------|-----------|--------------|
| Query Match | 9.0%; | Score 43.8; | DB 4; | Length 1803; |
| Best Local Similarity | 50.7%; | Pred. No. 4.9; | | |
| Matches 105; Conservative | 0; | Mismatches 102; | Indels 0; | Gaps 0 |

Oy 48 AATATTTTACGTGATCTCTACAAGTGTATATATGTTGGTGAATCTATTATTG 107
 Db 129 AGAATATTTTATTTCTTCCACAGCAAGATGTTGGTGGAGGTGAATACATCTTTTAT 188
 Oy 108 AGAGTATTTTAAACACCTTGAATCTAAATCTATATTAATAATATTTCTATCTTAA 167
 Db 189 CTAGTCTTAAATATATATTTTAAAGGAAAAATTTATGAAGAATTTAGTTAACTCT 248
 Oy 168 AAGCACATTTTCGTGGCTAAGGCATTTACGCTGATATATCTGAAACTCATGTCGCC 227

Db 249 GACTGACATTTTGGATTAAGAATTGGCTTTGTTGATAGACATCATTAAGAAAT 308

Qy 228 ACTAAATTCCTTCTTAACAGCGTTCTGCT 254

Db 309 ATTGAATAGACATATTGTGATCTTT 335

Search completed: December 28, 2005, 11:32:43
Job time : 825 secs

This Page Blank (uspto)

| | | | | | | |
|----|------|-----|--------|---|---------------------|----------------------|
| 1 | 41.8 | 8.6 | 177175 | 7 | US-11-121-086-79 | Sequence 79, App1 |
| 2 | 41.6 | 8.6 | 6801 | 6 | US-10-240-708-62 | Sequence 62, App1 |
| 3 | 38.8 | 8.0 | 148220 | 7 | US-11-121-086-90 | Sequence 90, App1 |
| 4 | 38 | 7.8 | 1303 | 6 | US-10-750-185-50224 | Sequence 50224, App1 |
| 5 | 38 | 7.8 | 95223 | 7 | US-11-117-187-18 | Sequence 188, App1 |
| 6 | 38 | 7.8 | 118936 | 7 | US-11-121-086-84 | Sequence 84, App1 |
| 7 | 37.4 | 7.7 | 684 | 6 | US-10-750-185-55780 | Sequence 55780, App1 |
| 8 | 37.4 | 7.7 | 171486 | 7 | US-11-121-086-105 | Sequence 105, App1 |
| 9 | 37 | 7.6 | 1150 | 6 | US-10-750-185-64497 | Sequence 64497, App1 |
| 10 | 37 | 7.6 | 207835 | 7 | US-11-121-086-39 | Sequence 39, App1 |
| 11 | 37 | 7.6 | 207835 | 7 | US-11-121-086-40 | Sequence 40, App1 |
| 12 | 36.6 | 7.5 | 201 | 6 | US-10-995-561-73227 | Sequence 73227, App1 |
| 13 | 36.6 | 7.5 | 1048 | 6 | US-10-750-185-30231 | Sequence 30231, App1 |
| 14 | 36.2 | 7.5 | 1588 | 6 | US-10-750-185-33457 | Sequence 33457, App1 |
| 15 | 36.2 | 7.5 | 49979 | 6 | US-10-995-561-13443 | Sequence 13443, App1 |
| 16 | 36 | 7.4 | 2248 | 6 | US-10-750-185-52939 | Sequence 52939, App1 |
| 17 | 36 | 7.4 | 165158 | 6 | US-10-995-561-13304 | Sequence 13304, App1 |
| 18 | 35.8 | 7.4 | 10467 | 6 | US-10-240-708-2 | Sequence 2, App1 |
| 19 | 35.8 | 7.4 | 645179 | 6 | US-10-995-561-13293 | Sequence 13293, App1 |
| 20 | 35.6 | 7.3 | 3499 | 6 | US-10-750-185-46381 | Sequence 46381, App1 |
| 21 | 35.6 | 7.3 | 167375 | 7 | US-11-121-086-93 | Sequence 93, App1 |
| 22 | 35.4 | 7.3 | 1579 | 6 | US-10-750-185-36438 | Sequence 36438, App1 |
| 23 | 35.4 | 7.3 | 19513 | 6 | US-10-240-708-40 | Sequence 40, App1 |

| | | | | | | |
|------|------|-----|---------|---|---------------------|--------------------|
| C 24 | 35.2 | 7.3 | 3350 | 6 | US-10-793-626-3453 | Sequence 3453, Ap |
| C 25 | 35.2 | 7.3 | 151870 | 6 | US-10-995-561-13199 | Sequence 13199, A |
| C 26 | 35 | 7.2 | 887 | 6 | US-10-750-185-52828 | Sequence 52828, A |
| C 27 | 35 | 7.2 | 203867 | 7 | US-11-121-086-50 | Sequence 50, Appl |
| C 28 | 34.8 | 7.2 | 6866 | 6 | US-10-240-708-20 | Sequence 20, Appl |
| C 29 | 34.8 | 7.2 | 7673 | 6 | US-10-516-768-15 | Sequence 15, Appl |
| C 30 | 34.6 | 7.1 | 1654 | 6 | US-10-750-185-50592 | Sequence 50592, A |
| C 31 | 34.6 | 7.1 | 1825 | 6 | US-10-750-185-58628 | Sequence 58628, A |
| C 32 | 34.6 | 7.1 | 137935 | 6 | US-10-995-561-13278 | Sequence 13278, A |
| C 33 | 34.6 | 7.1 | 215308 | 7 | US-11-121-086-77 | Sequence 77, Appl |
| C 34 | 34.4 | 7.1 | 2007 | 6 | US-10-995-561-42981 | Sequence 42981, A |
| C 35 | 34.4 | 7.1 | 2071 | 6 | US-10-750-185-58341 | Sequence 58341, A |
| C 36 | 34.4 | 7.1 | 1082114 | 7 | US-11-117-187-211 | Sequence 211, Appl |
| C 37 | 34.2 | 7.1 | 1755 | 6 | US-10-750-185-24875 | Sequence 24875, A |
| C 38 | 34.2 | 7.1 | 5629 | 6 | US-10-276-2334-13 | Sequence 13, Appl |
| C 39 | 34.2 | 7.1 | 403278 | 6 | US-10-995-561-13421 | Sequence 13421, A |
| C 40 | 34 | 7.0 | 1080 | 6 | US-10-750-185-54697 | Sequence 24697, A |
| C 41 | 34 | 7.0 | 2072 | 6 | US-10-750-185-58146 | Sequence 38146, A |
| C 42 | 34 | 7.0 | 2575 | 6 | US-10-750-185-44880 | Sequence 44880, A |
| C 43 | 34 | 7.0 | 1125000 | 6 | US-10-995-561-13286 | Sequence 13286, A |
| C 44 | 33.8 | 7.0 | 600 | 6 | US-10-750-185-5073 | Sequence 3073, Ap |
| C 45 | 33.8 | 7.0 | 600 | 6 | US-10-750-185-3340 | Sequence 3340, Ap |

ALIGNMENTS

```

RESULT 1
US-11-121-086-79
Sequence 79, Application US/11121086
Publication No. US20050266459A1
GENERAL INFORMATION:
APPLICANT: POULSEN, TIM S.
APPLICANT: NIELSEN, KRISTEN V.
TITLE OF INVENTION: NUCLEIC ACID PROBES A
FILE REFERENCE: 09138, 6000-0000
CURRENT APPLICATION NUMBER: US/11/121, 0866
CURRENT FILING DATE: 2005-05-04
PRIORITY APPLICATION NUMBER: 60/567,570
PRIORITY FILING DATE: 2004-05-04
NUMBER OF SEQ. ID NOS: 107
SOFTWARE: PatentIn version 3.3
SEQ. ID NO 79
LENGTH: 177175
TYPE: DNA
ORGANISM: Homo sapiens
US-11-121-086-79

```

Query Match 8.6%; Score 41.8; DB 7; Length 177175;

Best Local Similarity 51.3%; Pred. No. 1.2;
Matches 97; Conservative 0; Mismatches 92; Indels 0; Gaps 0

24 TCCITCGTTGCGTAGACAGTATGAAATATTTTACTGTGATACTTACAGTTGATATAT 83

Db 76610 TCCTGCTACATATACATTTTCAAAACAGCTGACTGTATGCTTAAATGTGATATAT 76669

84 GGTGTGTAATAATTGAGAGGTAATTAACACACCTTAGAACTAAACTTAA 143

Db 76670 TATTTCATGTAATTATATTCATTAATAAATTAATCTAAAAAGATTAAATAAAAGGATTTAG 767

144 TAAATAATATTTCTCTATCTTTAAAGGCACATATTACGTGGCTAAGGCAATTACAGCTG 203

Db 76730 ATTCTAGATCTCATGATATCTGTACACTTCTTTTAACCTCATAGAGTAATTCCTCTG 76731

QY 204 ATATACTGT 212

Db 76790 TTTTAAATT 767

RESULT 2

; Sequence 62, Application US/10240708

THE UNIVERSITY OF CHICAGO

```
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/240,708
; CURRENT FILING DATE: 2002-10-03
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 62
; LENGTH: 6801
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; US-10-240-708-62

Query Match      8.6%; Score 41.6; DB 6; Length 6801;
Best Local Similarity 44.1%; Pred. No. 0.52;
Matches 173; Conservative 0; Mismatches 219; Indels 0; Gaps 0;

QY 60 TGTGATCTTCAAGTGTGATATGTTGTGTGTAAGTCTTATTTTGAAGAGATTTTA 119
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5358 TTTGTAATTTAGATATTTTATTAAGGTTTGTAGTATTTAGATTTATTAATTAATTTTAG 5417

QY 120 ACAACCTTAAGACCTTAATTAATTAATTTCTCTAATCTTAAGACATAT 179
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5418 TTTATTAATAGAAATTAATTTTATTTAAATGATTAAGGATTTAAATGATTTT 5477

QY 180 ACGTGCTAAGGCAATTAAGCTGATATACGTAAACTGATGTCGCCACTAATCTTC 239
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5478 TCGAAGAGATGATGATGATGATTAATTAATGAATAATTTATTTATTTAGTATTA 5537

QY 240 TAAACAGCGTCTGTCTCTTCCAAAGGACTCGCAATATGCCATTTATCTGCGCAT 299
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5538 GGGAAATGTAATTAAGTATGATGATATTTTATTTTATTTTGAAGTTT 5597

QY 300 TTCCATTTATATCCCTATTTGGTATTTGATGTCGCCGTTAAATGACCGATTTGA 359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5598 TTGTAATTTTATTTTGTGTTATTAATTTTATTTTATTTTATTTTATTTTATTTT 5657

QY 360 ATCTCACTGTTGAGTTTGTCTTTGCTTCTTAAGGCTTCAATTTATCTAAAGC 419
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5658 TTTTATTTTATTTTTCGTTTCGTTTATTTTATTTTATTTTATTTTATTTTATTT 5717

QY 420 AAGTTTGTATTAATTCAAATACTTGTCTTT 451
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5718 TTTTATTTTATTTATTTATTTGATTTGCTTTGTTATTT 5749

RESULT 3
US-11-121-086-90/c
; Sequence 90, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
```

```
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 90
; LENGTH: 148220
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-11-121-086-90

Query Match      8.0%; Score 38.8; DB 7; Length 148220;
Best Local Similarity 52.5%; Pred. No. 6.1;
Matches 85; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 17 AAGTACTCTTCGTTGCGTAGACAGTATGAAAAATTTTACTGATCTTACAGATT 76
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 77435 AATCATCTCTTTATGCGTAGAATGATTTATATATATATATATATATATATATA 77376

QY 77 GATTAATGTTGTGTGTACCTTATTTTATTTGAGGATTTTAAACACCTTGAAGCTAA 136
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 77375 TATATATATATATATACCAATATATATATATATATATATATATATATATATAC 77316

QY 137 AACTTAATAATAATATTTCTCTATCTTTAAAGGACATAT 178
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 77315 ACACACACACACATATATATATATATATATATATATAT 77274

RESULT 4
US-10-750-185-50224
; Sequence 50224, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENBERG, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50224
; LENGTH: 1303
; TYPE: DNA
; ORGANISM: Bovine
; US-10-750-185-50224

Query Match      7.8%; Score 38; DB 6; Length 1303;
Best Local Similarity 57.6%; Pred. No. 2.3;
Matches 68; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 351 ACCGATGATCTTCACTGTTGAGTTTGTCTTTGCTTCTTAAGGCTTCAATTT 410
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 741 ACTTATTCCTGTTCATTTAGCTTTGTCTCATTTCTTTTGTCTTCTTTC 800

QY 411 ATCTAAGCAAGTTTGTATTAATTCAAATFACCTTGTCTTCTCAATGACTGAACCT 468
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 801 AATTAATTAATTTTATTAATTCATCTTATTCCTTTTATTAATTCATTCACCAT 858

RESULT 5
US-11-117-187-188/c
; Sequence 188, Application US/11117187
; Publication No. US20050266560A1
; GENERAL INFORMATION:
; APPLICANT: PREUSS, DAPHNE
; APPLICANT: COPEHAEVER, GREGORY
; TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS
```

FILE REFERENCE: ARCD:309US
CURRENT APPLICATION NUMBER: US/11/117,187
CURRENT FILING DATE: 2005-04-28
PRIOR APPLICATION NUMBER: US/09/531,120
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: 60/125,219
PRIOR FILING DATE: 1999-03-18
NUMBER OF SEQ ID NOS: 212
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 188
LENGTH: 95223
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: modified base
LOCATION: (2720)..(72695)
OTHER INFORMATION: N = A, C G, or T/U
US-11-117-187-188

Query Match
Best Local Similarity 7.8%; Score 38; DB 7; Length 95223;
Best Local Similarity 50.5%; Pred. No. 8.3;
Matches 92; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 36 TAGACAGTATGAATATTTTACTGTGATCTGATCAAGTTGATATATGCTGTGTGA 95
DB 24588 TATACATATCAAAATAGTGTATCTTCGCTGACACATCTAAATGTGTAATATA 24529
QY 96 CTTATTTATTTAGAGTATTTTAAACACACCTTGAACCTTAAATTAATTAATTT 155
DB 24528 CACAGTACTATATTCATATGTTTACGAGTTAAATCTTAAACCAATTAAGATTTAT 24469
QY 156 TCTTATCTTTTAAAGGCACATATTAAGTGGCTAAGGCAATTAAGCTGATATCTGTA 215
DB 24468 TTTGTAATTTTATATTAAGTTTCTGACTAAATAAATATATCATGTAGTGTACC 24409
QY 216 AC 217
DB 24408 AC 24407

RESULT 6
US-11-121-086-84
Sequence 84, Application US/11/121086
Publication No. US20050266459A1
GENERAL INFORMATION:
APPLICANT: POULSEN, TIM S.
APPLICANT: NIELSEN, KIRSTEN V.
TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
FILE REFERENCE: 09138,6000-00000
CURRENT APPLICATION NUMBER: US/11/121,086
CURRENT FILING DATE: 2005-05-04
PRIOR APPLICATION NUMBER: 60/567,570
PRIOR FILING DATE: 2004-05-04
NUMBER OF SEQ ID NOS: 107
SOFTWARE: PatentIn version 3.3
SEQ ID NO 84
LENGTH: 118996
TYPE: DNA
ORGANISM: Homo sapiens
US-11-121-086-84

Query Match
Best Local Similarity 7.8%; Score 38; DB 7; Length 118996;
Best Local Similarity 45.8%; Pred. No. 8.9;
Matches 131; Conservative 0; Mismatches 155; Indels 0; Gaps 0;
QY 51 TATTTTACTGTGATCTTACAGTTGATATATGCTGTGTGTAATTTTATTTGAGA 110
DB 9710 TCTTTCTAATTCATCTTCTAATTTTAAAGTATGCTGTGCTTATTTTACCTTGTAA 9769
QY 111 GGTATTTTAAACACCTTGAATAAACTTAAATAAATATTTCTTATCTTTAAAG 170
DB 9770 GCCGTTCAATCTTTTCTGAATATAAACAAGATACATTAAGGTGTAATAAGATCTAT 9829

QY 171 GCACATATTAGTGCTTGAAGCAATTACGCTGATATATCTGTAATCTATGTCGCACT 230
DB 9830 GTTTATATTTCATCTACTCATCTATGAAACCTGAAGCTAGCGCTAATTAATCAATG 9889
QY 231 AAATCTCTTAAACAGCGTGTCTGCTCTTCTTCCAAAGGAGCTCGATATGCACTTTAT 290
DB 9890 ATGCTCTCATATGACCTTTTCAAGTCAATGAAGTGTGCTCCATGATTAATTAATGAG 9949
QY 291 CTGTGCAATTTCCATTTATATATTCCTTATTTGGTATTTGATGAG 336
DB 9950 CTGAATAAATCTTCATGCTGCTAGTACCTCATAGCATATGATGTGG 9995

RESULT 7
US-10-750-185-55780
Sequence 55780, Application US/10750185
Publication No. US20050260603A1
GENERAL INFORMATION:
APPLICANT: MMT GENOMICS, INC.
APPLICANT: DENISE, Sue K.
APPLICANT: KERR, Richard
APPLICANT: ROSENFELD, David
APPLICANT: HOLM, Tom
APPLICANT: BATES, Stephen
APPLICANT: FANTIN, Dennis
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
FILE REFERENCE: Mm1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
CURRENT FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 3.1
SEQ ID NO 55780
LENGTH: 684
TYPE: DNA
ORGANISM: Bovine
US-10-750-185-55780

Query Match
Best Local Similarity 7.7%; Score 37.4; DB 6; Length 684;
Best Local Similarity 53.8%; Pred. No. 2.7;
Matches 77; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 54 TTTTACTGTGATCTTACAGTTGATATATGCTGTGTGTAATTTTATTTTGAAGCT 113
DB 504 TTTTCAATATATGATATATGAAGAAAGACATGTAATTTAAATTAATTTTGAAGCT 563
QY 114 ATTTAACAACCTTGAAGCAATTAATTAATTAATTTCTTATCTTTAAAGCA 173
DB 564 TTTGATATGAGCTTTTAAATGTAATGCAATATATATATATATTAATTAATTAATAA 623
QY 174 CATATTAGTGCTTGAAGCAATT 196
DB 624 TTTATTAATTTTAAATTAATTT 646

RESULT 8
US-11-121-086-105
Sequence 105, Application US/11/121086
Publication No. US20050266459A1
GENERAL INFORMATION:
APPLICANT: POULSEN, TIM S.
APPLICANT: NIELSEN, KIRSTEN V.
TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
FILE REFERENCE: 09138,6000-00000
CURRENT APPLICATION NUMBER: US/11/121,086
CURRENT FILING DATE: 2005-05-04
PRIOR APPLICATION NUMBER: 60/567,570
PRIOR FILING DATE: 2004-05-04
NUMBER OF SEQ ID NOS: 107
SOFTWARE: PatentIn version 3.3
SEQ ID NO 105
LENGTH: 171486

```

; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-105

```

| | | | | |
|-----------------------|--------|---------------|-------|----------------------------------|
| Query Match | 7.7%; | Score 37.4; | DB 7; | Length 171486; |
| Best Local Similarity | 59.0%; | Pred. No. 14; | | |
| Matches | 82; | Conservative | 0; | Mismatches 56; Indels 1; Gaps 1; |

| Qy | Db | Qy | Db |
|---|---|---|--|
| 43 | 17286 | 102 | 17346 |
| TATGAAATATTTTCTAGTGCATCTTCAAGTGTGATATGGTGTGCGTA-ACCTATT | TATGAAATATTTTCTAGTGCATCTTCAAGTGTGATATGGTGTGCGTA-ACCTATT | TATTTGGAGGTAATTTTACACACTTGAAGACTTAAATCTTAAATTAATTTCTCTA | TATTTAGCATATATATGAAATATATTTTATTTATTTATTTATTTATTTATTTATTA |
| 17345 | 17405 | 161 | 17405 |
| TATGAAATATTTTCTAGTGCATCTTCAAGTGTGATATGGTGTGCGTA-ACCTATT | TATGAAATATTTTCTAGTGCATCTTCAAGTGTGATATGGTGTGCGTA-ACCTATT | TATTTGGAGGTAATTTTACACACTTGAAGACTTAAATCTTAAATTAATTTCTCTA | TATTTAGCATATATATGAAATATATTTTATTTATTTATTTATTTATTTATTTATTA |

| | | | | | | | | | | |
|----|-------|----|-----|----|----|-----|----|-----|----|-------|
| Qy | 162 | TC | TT | AA | GG | CA | CA | TAT | TA | 180 |
| | | | | | | | | | | |
| Db | 17406 | TT | TAT | AG | AT | TAT | AT | AT | TA | 17424 |

```

1  RESULT 9
2  US-10-750-185-64497
3  Sequence 64497, Application US/10750185
4  Publication NO. US20050260603A1
5
6  GENERAL INFORMATION:
7  APPLICANT: MMI GENOMICS, INC.
8  APPLICANT: DENISE, Sue K.
9  APPLICANT: KERR, Richard
10 APPLICANT: ROSENFELD, David
11 APPLICANT: HOLM, Tom
12 APPLICANT: BATES, Stephen
13 APPLICANT: FANTIN, Dennis
14 APPLICANT: TITLE OF INVENTION: COMPOSITIONS FOR INFE
15 FILE REFERENCE: MM1100-2
16 CURRENT APPLICATION NUMBER: US/10/750,185
17 CURRENT FILING DATE: 2003-12-31
18 PRIOR APPLICATION NUMBER: US 60/437,482
19 PRIOR FILING DATE: 2002-12-31
20 NUMBER OF SEQ ID NOS: 64922
21 SOFTWARE: PatentIN version 3.1
22
23 SEQ ID NO 64497
24 LENGTH: 1150
25
26 TYPE: DNA
27
28 ORGANISM: Bovine 19866881242828
29
30 US-10-750-185-64497

```

| | | | | |
|-----------------------|--------|--|-------|----------------|
| Query Match | 7.6%; | Score 37; | DB 6; | Length 1150; |
| Best Local Similarity | 58.7%; | Pred. No. 3.9; | 45; | Gaps 0; |
| Matches | 64; | Conservative | 0; | Mismatches 45; |
| | | | | Indels 0; |
| | | | | Gaps 0; |
| Qy | 32 | TCGCTAGACAGTANGAAATATTTTACGTGACTCTCAAGTTGATATATGTTGCT | 91 | |
| | | | | |
| Db | 266 | TACATTTAAATGACATATTTAATTAATAGATATCAAAATTCATTTACTTTGTTT | 325 | |
| | | | | |
| Qy | 92 | GTAACATATTTATTTGAGAGTATTTTAAACACCTTAGAAGTAAACT | 140 | |
| | | | | |
| Db | 326 | ATAATTTGTTTTCTGAAGATATTTTGAAGGAGTATCAAGAAAAAATPAT | 374 | |
| | | | | |

RESULT 10
US-11-121-086-39
Sequence 39, Application US/11121086
Publication No. US20050266459A1
GENERAL INFORMATION:
APPLICANT: NIELSEN, TIM S.
APPLICANT: NIELSEN, KIRSTEN V.
TITLE OR INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
FILE REFERENCE: 09138.6000-00000
CURRENT APPLICATION NUMBER: US/11/121,086
CURRENT FILING DATE: 2005-05-04
PRIORITY APPLICATION NUMBER: 60/567,570
PRIORITY FILING DATE: 2004-05-04

```

; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: Patentin version 3.3.3
; SEQ ID NO 39
; LENGTH: 207835
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-39

```

| | | | | |
|---------------------------|--------|-----------------|-----------|----------------|
| Query Match | 7.6%; | Score 37; | DB 7; | Length 207835; |
| Best Local Similarity | 48.4%; | Pred. No. 18; | | |
| Matches 103; Conservative | 0; | Mismatches 110; | Indels 0; | Gaps 0; |

| | | | |
|----|--------|---|--------|
| QY | 78 | ATATATGGTGTGCTACTTATTTATTTGAGAGTATTTACACACCTTACACATTTAA | 159 |
| Db | 159868 | ATGTTAAATTTAAATTTAAATTTAGATACATTTAAATTTGTCAAAGTCTCTCTTAA | 159947 |
| QY | 138 | ACTTATATTAATTAATTTCTCTACTCTTAAAGCACAATTACGTGCTAAAGCAATTA | 197 |
| Db | 159948 | GCTTCAAGAAAATTAATATCAATCTCCTTTAAATGAACAATATATGATGAATTCGTCAAAA | 200007 |

| | | |
|--------------------------------|---------------------------------|--------|
| 258 | TTTCCAGGCACTCCGAATATGCCACTATTAT | 290 |
| 200068 | TTAAAGACAGATTTAGAAATACACTAATTTT | 200100 |
| Db | | |
| 200008 | ATATTATATTTGTGTAACTATTACCAAGCA | 200068 |
| ATATTATATTTGTGTAACTATTACCAAGCA | TTCAATTTACATCTTGTGTCTG | |
| Db | | |
| Qy | | |

```

RESULT 11
US-11-121-086-40
; Sequence 40, Application US/11121086
; Publication No. US2005026459A1
GENERAL INFORMATION:
APPLICANT: BOULSEN, TIM S.
INVENTOR: NIELSEN, KRISTEN V.
TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
FILE REFERENCE: 09136.6000-0000
CURRENT APPLICATION NUMBER: US/11/121,086
CURRENT FILING DATE: 2005-05-04
PRIOR APPLICATION NUMBER: 60/567,570
PRIOR FILING DATE: 2004-05-04
NUMBER OF SEQ ID NOS: 107
SOFTWARE: PatentIn version 3.3
SEQ ID NO 40
LENGTH: 207835
TYPE: DNA
ORGANISM: Homo sapiens
US-11-121-086-40

```

| | | | | |
|---------------------------|--------|-----------------|-----------|----------------|
| Query Match | 7.6% | Score 37; | DB 7; | Length 207835; |
| Best Local Similarity | 48.4%; | Pred. No. 18; | | |
| Matches 103; Conservative | 0; | Mismatches 110; | Indels 0; | Gaps 0 |

Qy 78 ATATTAAGGTTGGTGTGAACCTTAATTATTTGGAAGGTAITTTACACACCTTAAACATATAA 197
 Db 199888 ATGTAAATTTAATAAAATTTTAAATTTTGATACATTAAAAATGTCMAAGTCCCTCTTAA 199947
 Qy 138 ACCTTAATAATAAATTTTCTCATCTTTAAAGGCACATTAACGCGGTCTAAGSCATTA 197
 Db 199948 GCTTCAAGAAAATAAATCAATCATCTCTTAAATGAACAAATTTATGATGAATTCGTAAAA 200067
 Qy 198 CAGCTGATATACTGTAAAACCTCATGTGGCACAATAATCTTCTTACACGGGTTCTGTCTC 257
 Db 200008 ATATTATTTATGGTTTAACTCATTTTACCAAGACATTTCAAAATTTACATCTTGTGTCTG 200067

| Qy | 258 | TTTCCAGGGA | CTCCGAAT | ATGCCA | CTATT | TTAT | 290 |
|----|--------|------------|----------|--------|--------|------|--------|
| | | | | | | | |
| Db | 200068 | TTAAAGCAG | ATTAGAA | TACAC | TAATTT | TAT | 200100 |

RESULT 12
US-10-995-561-73227/c

| | Query Match | Score | DB | Length |
|-----------------------|-------------|--------------|----|---------------------------------|
| Best Local Similarity | 7.5% | 36.6 | 6 | 1048 |
| Matches | 96 | Conservative | 0 | Mismatches 99; Indels 0; Gaps 0 |

| | Query | DB |
|-----|--|-----|
| 283 | CAATTATCTGCGCATTTCCCAATTATATCCCATGGGATTTGATGTCGCCGTT | 342 |
| 60 | CTCGCTCTTAGGCGCATTCCTCAGTCATCTTAACATTTTTCATATCTAATTAATGACA | 119 |
| 343 | AAATGTCACCGATGGAATCTTCACTGTGTCGAGTTTGTCTTTGCTTCTCTAAAGCTC | 402 |
| 120 | ACTCATTTGCTCTAATTAATTAACATCTGTGCTTTTTCGACAAATTTTGTTCATTAACATTA | 179 |

```

/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
/ TITLE OF INVENTION: DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001559
/ CURRENT APPLICATION NUMBER: US/10/995,561
/ CURRENT FILING DATE: 2004-11-24
/ NUMBER OF SEQ ID NOS: 85702
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 13443
/
/ LENGTH: 49979
/
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(49979)
/

```

; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13443

Query Match 7.5%; Score 36.2; DB 6; Length 49979;

Best Local Similarity 50.3%; Pred. No. 19;
Matches 86; Conservative 1; Mismatches 84; Indels 0; Gaps 0;

QY 9 AAGAGGAAAGTGAATCTCTCGTGGCTAGACAGTATGAAAATATTTTACTGTGATCT 68
DB 43936 AACCAAGAGAAATACTCATTTATATCTAAACAATAGTCTTATGATMAAGGTTAGT 43937
QY 69 TACAAGTATATATGCTGTGTGTAAGTATTTTATTTGAGAGTATTTTAAACACCTT 128
DB 43936 TTAAATGGATACAAAATTTGCTGTGTAATAAAGTGTTCAAAATACATTCTATAGGTA 43877
QY 129 AGAAGTAAAGCTTAAATAAATATTTCTATCTTTTAAAGGACATATT 179
DB 43876 GAGACTATGTCTTAGTAAAGAGCAGTTATCTATTTATCAAAATATCTATT 43826

Search completed: December 28, 2005, 11:37:47
Job time : 298 secs

November 2005

Published_Applications Nucleic Acid and Published_Applications Amino Acid database searches now generate two sets of results each. The Published_Applications databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Newly published applications will appear in the Published_Applications_New databases; older published applications make up the Published_Applications_Main databases.

- * Searches run against Nucleic Acid Published_Applications produce two sets of results, with the extensions **.rnpbm** (Published_Applications_NA_Main) and **.rnpbn** (Published_Applications_NA_New).
- * Searches run against Amino Acid Published_Applications produce two sets of results, with the extensions **.rapbm** (Published_Applications_AA_Main) and **.rapbn** (Published_Applications_AA_New).

This Page Blank (uspto)